Perception of sound symbolism in mimetic stimuli:
The voicing contrast in Japanese and English

By

Kotoko Nakata

Submitted to the graduate degree program in Linguistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

____
Chairperson: Dr. Joan Sereno

____
Dr. Allard Jongman

____
Dr. Utako Minai

Date Defended: August 23rd, 2013

The Thesis Committee for Kotoko Nakata
certifies that this is the approved version of the following thesis:

Perception of sound symbolism in mimetic stimuli:
The voicing contrast in Japanese and English

____

Chairperson: Dr. Joan Sereno

Date approved: August 23, 2013

Abstract

Sound symbolism is a concept in which the sound of a word and the meaning of the word are systematically related. Some aspects of sound symbolism have been found to be language-specific and some to be cross-linguistic. The current study investigated sound symbolism in Japanese using Japanese mimetic non-words. In this study, voicing of consonants was examined and vowel influence was controlled. We examined whether the voicing contrast in consonants (/t, k, s/ vs. /d, g, z/) affects the perception in both Japanese native speakers and English native speakers who had no knowledge of Japanese. Two additional manipulations were also included. First, stimuli were evaluated on 4 different dimensions including both size (big-small) and shape (round-spiky) dimensions as well as evaluative dimensions (good-bad, graceful-clumsy), in order to examine the generality of the sound symbolism. Second, voicing was manipulated, creating a continuum from voiced to voiceless endpoints, in order to examine the categorical nature of the perception. In the current study, both Japanese and English speakers tended to associate voiced sounds with largeness, badness, and clumsiness and voiceless sounds with smallness, goodness, and gracefulness. In addition, the current study found, for a shape dimension, a tendency in English speakers to associate voiced stop consonants with roundness and voiceless stops with spikiness. This tendency was observed when the stimuli consisted of only stops and a vowel, but not when they also contained fricative consonants. Sound symbolism in Japanese and English is discussed.

Acknowledgements

I am deeply thankful to my committee chair, Dr. Joan Sereno whose encouragement, guidance and support from the initial to the final level enabled me to complete the thesis. Without her guidance and persistent help this thesis would not have been possible. I would also like to express my gratitude to my committee members, Dr. Allard Jongman and Dr. Utako Minai, in addition to those who supported me in any respect during the completion of this study.

Table of contents

Perception of sound symbolism in mimetic stimuli: The voicing contrast in Japanese and English

1. Introduction

Sound symbolism is a concept in which the sound of a word and the meaning of the word are systematically related. Although it is generally acknowledged that the relation between sound and meaning is arbitrary in the linguistic theory by de Saussure (1915), a number of studies have been conducted to investigate this relationship in many languages since the 1920's. Some aspects of sound symbolism have been found to be language-specific and some to be cross-linguistic. In other words, they can be either specific to certain languages or shared across different languages. The current paper first examines whether the voicing contrast in Japanese affects two types of sound symbolism, size symbolism (big-small dimensions) and shape symbolism (round-spiky dimension) by conducting a perception experiment with native speakers of Japanese and native speakers of English who have no knowledge of Japanese. Second, evaluative sound symbolism (good-bad dimension and graceful-clumsy dimension) was examined in the similar manner.

2. Sound symbolism

Sound symbolism is when the sound of a word and the meaning of the word are systematically related. Despite a fundamental assumption that the relation between sound and meaning is arbitrary since the time of de Saussure (1915), recent evidence seems to support some version of sound symbolism that constrains mapping of meaning onto sounds.

2.1. Size symbolism

One of the major findings from Sapir (1929) is size symbolism. He found that given two non-words "mil" and "mal" and two tables of different size (small and large), English speakers tend to associate "mil" with a small table and "mal" with a large table. He claimed that English speakers tend to associate the vowel /i/ with smallness and the vowel /a/ with largeness. Ultan (1978) claimed that both height and backness of vowels also affect the size image. This size symbolism has been found in other languages such as Japanese (Hamano, 1998), Chinese, and Thai (Huang et al. 1969). The size symbolism has been found not only in vowels but also in consonants. Hamano(1998) found the size symbolism in voicing of initial obstruents in Japanese. She found that voiceless obstruents /p, t, s, k/ tend to indicate smallness/lightness in Japanese mimetic words as shown in examples 1 and 3 whereas voiced obstruents /b, d, z, g/ tend to indicate largeness/heaviness as shown in examples 2 and 4.

1. *<u>ton-ton</u>* *taiko-o* *tataku* *oto*
   drum-ACC hit sound
   'the (light) rhythmic sounds of beating small drums'

2. *<u>don-don</u>* *taiko-o* *tataku* *oto*
   drum-ACC hit sound
   'the (heavy) rhythmic sounds of beating large drums'

3. *Supurinkuraa-no mizu-ga* ***patto*** *kakatta.*
sprinkler-GEN water-NOM was sprayed on
'A splash from the sprinkler sprayed on me.'

4. *Mizu-o* ***batto*** *kaker-are-ta.*
Water-ACC pour-PASS-PAST
'He threw (a bucket of) water (and I was drenched).' (Hamano 1998, p.83-85)

Shinohara & Kawahara (to appear) tested size symbolism by investigating how the images of size (small or large) are affected by phonetic factors. They explored exactly which phonetic factors, the height and the backness of vowels, and voicing in obstruents, determine the image of size in four languages; Chinese, English, Japanese, and Korean. They had three research questions; (i) Whether the size symbolism holds across languages (ii) Which phonetic dimensions determine the image of size, albeit to different degrees across languages (iii) Whether phonetic grounding (e.g. opening of the mouth) affects sound symbolism.

There were four subject groups; speakers of Chinese (N=20), English (N=22), Japanese (N=42), and Korean (N=19). The stimuli were all non-words with VCVC form, in which the two vowels and the two consonants were identical. The vowels were either /i, e, a, o, u/ and the consonants were one of voiced obstruents /b, d, g, z/ or voiceless obstruents /p, t, k, s/. All the four target languages have the 5 vowels and the 8 consonants. The stimuli consisted of 40 disyllabic non-words in all the target languages (e.g. /ibib/). The subjects were instructed to listen to an exotic language and asked to rate the size of each words in the exotic language on a scale of 1-4 (1=very small, 2=relatively small, 3=relatively big, 4=very big).

Shinohara & Kawahara (to appear) found some consistent patterns across the target languages. They found that /i/ is rated as smaller than other vowels and /a/ and /o/ are rated as larger than other vowels. Thus, they answer the research question (i) by concluding that the size symbolism in vowels held across all languages. As for the phonetic factors that affect the size sound symbolism, they found effects in all three dimensions; the height of the vowels, the backness of the vowels, and the voicing of obstruents. First, there was a trend in which the lower the vowel, the larger the image (Figure 1). In all the languages, low vowels evoked larger images than high vowels. When comparing mid vowels and high vowels, mid vowels evoked larger images in all the target languages except Japanese. Shinohara & Kawahara (to appear) did not have a good explanation for this Japanese pattern found in this experiment.

Second, Shinohara & Kawahara (to appear) found that back vowels evoked significantly larger images than front vowels in all the target languages (Figure 2). Third, voiced obstruents /b, d, g, z/ evoked larger images than voiceless obstruents /p, t, k, s/ in all the target languages except Korean which showed a slightly reverse result (Figure 3). Therefore, they answered question (ii) by concluding that there are three phonetic factors that affect the size symbolism with a few exceptions; ①low vowels are associated with larger images than high vowels ②back vowels are associated with larger images than front vowels ③ voiced obstruents are associated with larger images than voiceless images.

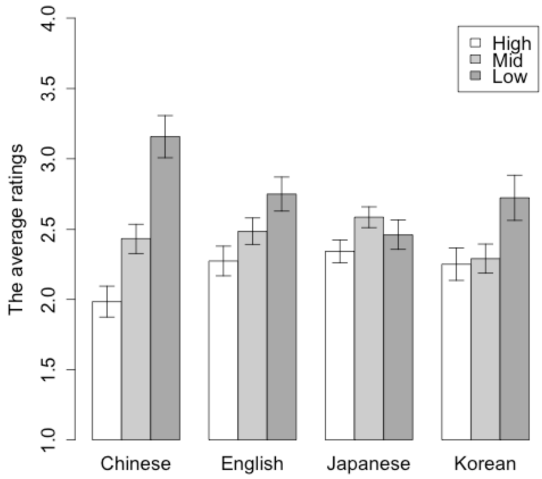

Figure 1: The effect of vowel height from Shinohara & Kawahara (to appear)

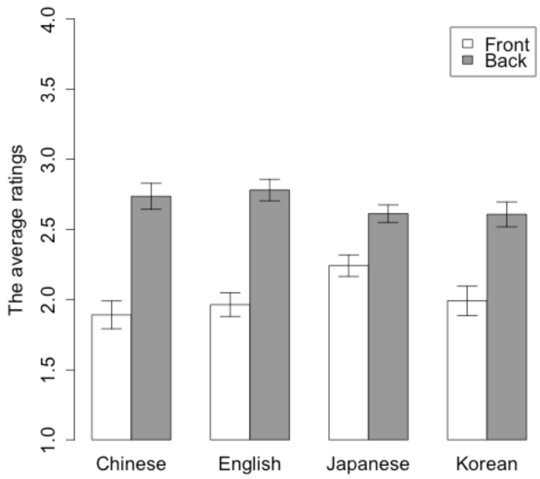

Figure 2: The effect of vowel frontness from Shinohara & Kawahara (to appear)

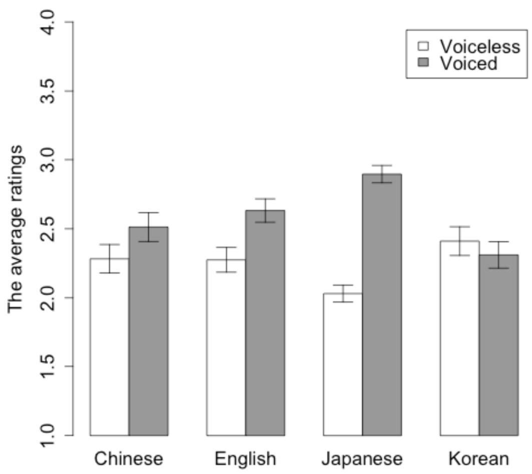

Figure 3: The effect of obstruent voicing from Shinohara & Kawahara (to appear)

They attempted to answer question (iii) whether phonetic grounding affects the size symbolism, by addressing articulatory and acoustic explanations for each of the three phonetic factors mentioned above. In terms of articulatory explanations, they claimed that the lower the vowel, the wider the aperture, and back vowels have a larger sub-oral cavity in front of the tongue. They suggested that this articulation explanation might lead to the sensation of larger images in lower and back vowels; the larger the aperture and the sub-oral cavity in the front of the tongue, the larger the image. In terms of acoustic explanations, they found F2 to be a good predictor of the size of the image. The judged size of five vowels roughly followed the order of /i/ < /e/ < /a/ < /u/ < /o/, and their F2 values followed the reverse order of /i/ > /e/ > /a/ > /u/ > /o/. They suggested that F2 inversely correlates with the size of images at least in their experiment.

The effect of voicing in obstruents could also be explained articulatorily. In order to pronounce voiced obstruents, speakers need to expand their oral cavities by lowering the larynx and making the air pressure in the oral cavity lower than the air pressure in the sub-glottal cavity. Thus, they suggested that this articulatory expansion of the oral cavity might lead to the sensation of larger images in voiced obstruents.

To sum up, Shinohara & Kawahara (to appear) concluded that size symbolism held across languages and the three phonetic factors (the height and the backness of vowels and the voicing in obstruents) affect sound symbolism. They also suggested that articulatory/acoustic explanations can be applied to the size symbolism for each of the three phonetic factors.

2.2. Shape symbolism

Another significant finding is the shape symbolism identified by Kohler (1929). He found that when presented with two non-words, “baluma” and “takete”, and two types of shapes, a curvy round shape and a spiky angular shape, one tends to associate “baluma” with the curvy round shape and “takete” with the spiky angular shape (Figure 4). They suggested that English speakers tend to associate continuant strings with curvy shapes and stop strings with spiky shapes. However, there are two related concerns need to be mentioned. First, the continuant stimulus “baluma” contains /b/ which is a stop consonant. Second, the stimuli used in this experiment may have had other differences other than the stop/continuant contrast that have been found to affect sound symbolism. For example, in recent studies as mentioned earlier, voicing contrasts in consonants and height/frontness differences in vowels have affected the sound symbolism.

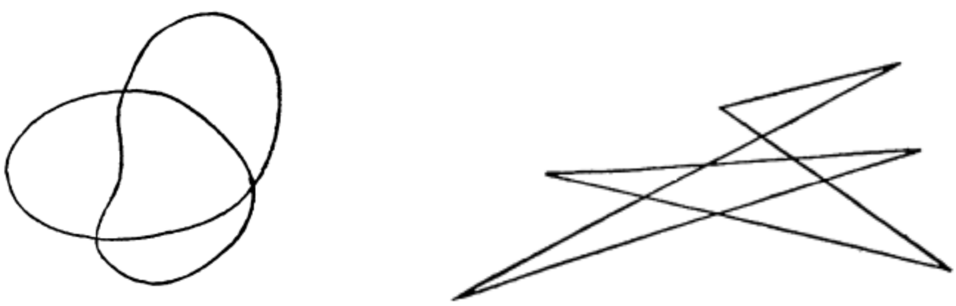

Figure 4: Reproductions of original “baluma” and “takete” stimuli from Kohler’s (1929)

Westbury (2005) replicated and extended Kohler’s claims by using both non-words and real words as stimuli. He hypothesized that if subjects really do associate continuant strings with curvy shapes and stop strings with spiky shapes, curvy shapes will facilitate the identification of all-continuant strings whereas spiky shapes will facilitate the identification of all-stop strings.

Two parameters were manipulated in his experiment using a visual lexical decision task; the frame (curvy or spiky) in which the stimuli appeared, and the phonological structure of the letter strings (non-words and real words). All the stimuli were four-letters long. There were three types for non-words and real words; stop consonants only (e.g. “kide” and “toad”), continuant consonants only (e.g. “lole” and “moon”), and mixed stop and continuant consonants (e.g. “nuck” and “flag”). Each category (non-word and real word) had 20 of each letter string types. Therefore, in total, 60 non-words and 60 real words were used as stimuli. The frame stimuli consisted of white figures cut into a black background (Figure 5). The white figures were either curvy or spiky. Twenty frames of each shape were constructed.

Figure 5: Examples of the curvy frames and spiky frames from Westbury (2005)

Participants were 30 native English speakers. The subjects saw a string of letters in the middle of the frame. They were asked to decide as quickly and accurately as possible whether the string was a word. Each subject saw an equal number (120) of curvy and spiky frames, with the 40 frames being used three times per subject. An equal number (10 non-words and 10 real words) of each string type was shown within each frame type. Reaction times for correct decisions were measured.

The reaction time results for real words showed no frame × phonology interaction effect, both when analyzing by subjects and analyzing response times to words by items (Figure.6). On the other hand, the results for non-words showed a reliable in frame × phonology interaction effect that continuant non-words were recognized faster in curvy frames than spiky frames, and stop non-words were recognized faster in spiky frames than curvy frames (Figure.7). Therefore, the pattern of results seen for the non-words was consistent with the hypothesis. The predicted pattern of interaction was observed for non-words but not for real words. Since there were no semantic features in non-words in this experiment, they suggested that the sound symbolism effect must not be occurring at the level of semantic access, but rather at a prior stage of lexical access.

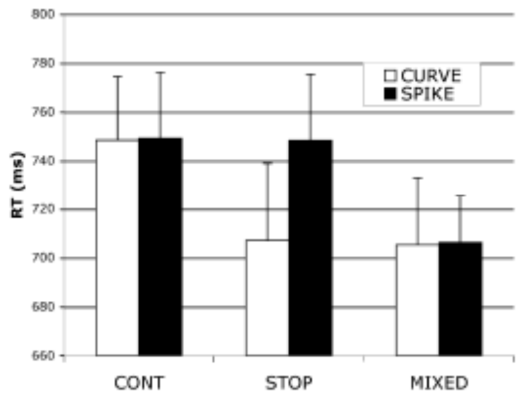

Figure 6: Reaction times to real words from Westbury (2005)

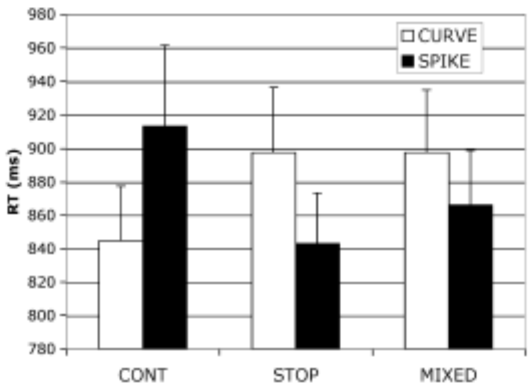

Figure 7: Reaction times to non-words from Westbury (2005)

Westbury (2005) designed another experiment to see how low in the linguistic hierarchy the effect could be demonstrated, by asking subjects to decide whether or not a single character was a letter, with numbers as foils. The same frames were used as frame stimuli as in the previous experiment. As for letter stimuli, there were three types; stop consonants (b, p, t, k), continuant consonants (f, m, n, r), and numbers from 2 to 9. The participants were 32 native English speakers. Each subject saw 4 practice stimuli and 64 test stimuli consisting of 32 letters and 32 numbers. They were asked to decide as quickly and accurately as possible if the character in the middle of the frame was a letter or not. The results showed that there was a reliable frame × character type interaction: the subjects took longer response times to correctly recognize a continuant consonant in a spiky frame than in a curvy frame, and longer to recognize a stop consonant in a curvy frame than in a spiky frame (Figure 8).

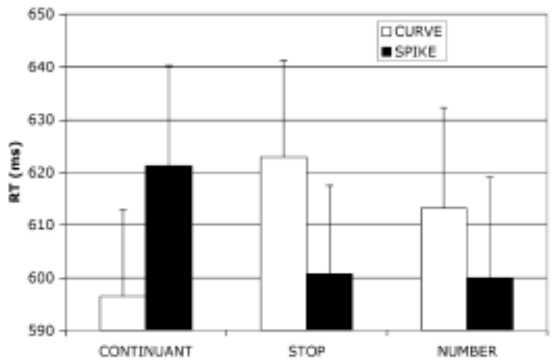

Figure 8: Reaction times in Experiment 2 from Westbury (2005)

These two experiments supported the existence of an interaction between visual form and phonology. They claimed that the relations between continuant consonants and curvy shapes and between stop consonants and spiky shapes have a psychological reality without informing subjects that they are being tested on the relation. However, it needs to be mentioned that vowels were not controlled in their experiments, suggesting that the effects might not only be due to the stop/continuant contrast in consonants but also due to vowel features used in the stimuli.

Aveyard (2012) recently found sound symbolism contained in plosive and non-plosive consonants for object recognition. Aveyard (2012) tested shape symbolism by conducting a learning experiment in which subjects were asked to learn word-picture pairs composed of unfamiliar words (Vietnamese) and unfamiliar objects (rectilinear objects and curvilinear objects). They hypothesized that if the relationship between sound and meaning is not arbitrary, sound information should enable faster learning of matching word-picture pairs. Each non-word stimulus (N=32) contained plosive phonemes /p, b, t, d, k, g/ (e.g. *kuh-der-pai*) or non-plosive phonemes /f, s, h, l, r, w/ (e.g. *fuh-lih-sai*). The non-word stimulus was paired with a rectilinear or curvilinear object.

The participants were college students of native Arabic speakers who had high proficiency in English at the American University of Sharjah in the United Arab Emirates. They were asked to learn word-picture pairs composed of "English non-words" and unfamiliar objects. First, they were presented with a word auditorily, then asked to click on one of two shapes appearing on the screen; one shape was a target (a rectilinear or curvilinear object) and the other was a distractor. 32 randomized trials were repeated three times (Round 1-3), and targets were paired with different distractor objects in each round. Feedback was provided after each trial to help participants improve their accuracy across three rounds of testing.

The results showed that the congruent word-picture pairs (plosives with rectilinear objects and non-plosives with curvilinear objects) evoked a greater advantage in object recognition than incongruent pairs (Figure 9). This provided evidence that plosive/non-plosive contrast in consonants evoke sound symbolism for object recognition at least for these second language learners.

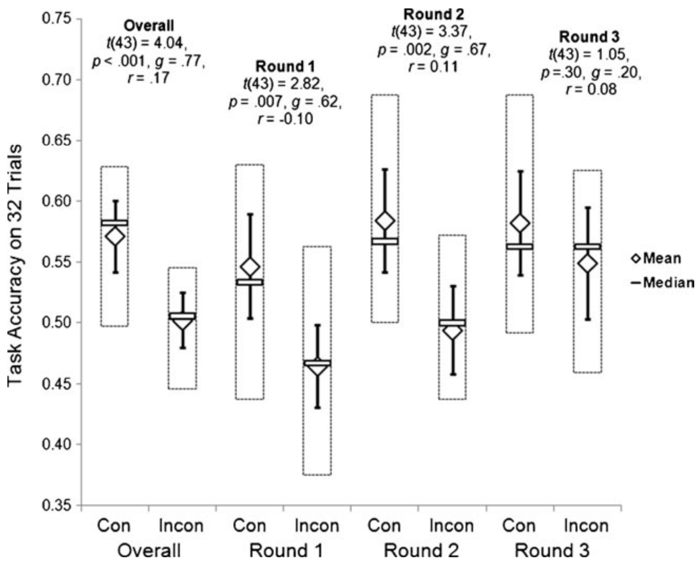

Figure 9: Overall accuracy across all rounds from Aveyard (2012)

2.3. Other sound symbolism

Besides these major examples of sound symbolism, size symbolism and shape symbolism, some other sound symbolism have also been found in recent studies. Parault & Parkinson (2008) conducted an experiment to assess 5th and 6th graders' knowledge of word meaning for English sound symbolic and non-sound symbolic words. There were two subject groups; native English speakers (N=22) and second language learners of English (N=24). The learners of English (ESL) attended a public elementary school in the mid-Atlantic region (their first languages were not mentioned in Parault & Parkinson, 2008). They were not proficient in English as defined by the IDEA Proficiency Test (IPT) even though the majority of them had a year or more exposure to English. They used two types of words as stimuli; 26 sound symbolic words and 26 non-sound symbolic words. All the stimuli (52 target words) were obsolete English words chosen from dictionaries by Grambs (1994), Halliwell (1850), and MacKay's (1879). None of the subjects had prior knowledge of the stimulus words. All 26 sound symbolic words contained one of 13 initial phoneme clusters that are found to be sound symbolic by Bloomfield (1933), Ciccotosto (1991) or Nuckolls (1999) as shown in the figures (Figure 10 & 11) below.

Sound symbol	Meaning
[cl-]	"a loud sound"
[cr-]	"noisy impact"
[dr-]	"watery, wet"
[fl-]	"moving light" or "movement in air"
[gl-]	"unmoving light"
[gr-]	"an angry manner"
[scr-]	"grating impact or sound"
[sl-]	"smoothly wet"
[sn-]	"creep"
[st-]	"walking movement"
[sw-]	"swift movement"
[tr-]	"effortful walking"
[tw-]	"a twisting or pinching motion"

Figure 10: 13 initial phoneme clusters from Parault & Parkinson (2008)

Stimulus word (real English words)	Definition
clicket	A clinking or jingling noise

craske	To smash
dreen	To remove all water
fluce	To flounce, to bounce
glisk	To sparkle
grame	Anger
scranch	To scratch , to rub
slive	To slip or skid down
snoove	To pry, to sneak, to creep
strake	To walk around
swaib	To move back and forth like a pendulum
traik	To wander without purpose
twage	To pinch, to squeeze

Figure 11: Examples of sound symbolic words containing one of the 13 initial phoneme clusters from Parault & Parkinson (2008)

Each stimulus word was presented both orally and visually to the subjects. The stimulus words were first presented with no surrounding context and the subjects were asked to guess what the phrase meant and generate a written definition for all the stimulus words. Next, the subjects were presented with the same list of words both orally and visually in a multiple-choice format (e.g. *glisk*: (a) to confuse, (b) to sparkle, (c) to lean (d) to pick). Both tasks were graded according to the 13 target clusters, their meanings, and the definitions of the stimulus words shown in the figures above. The generated definitions were scored on a 0-2 scale (0=incorrect definition or blank, 1=partially correct, indirect synonym, 2= completely correct, direct synonym). The multiple-choice (recognition task) was scored on a scale of 0-1 (0=incorrect answer, 1=correct answer).

The results showed that the sound symbolic words tend to yield more correct word definitions than the non-sound symbolic words in both tasks (guess phrase task and recognition task). There was no effect of language proficiency, meaning that both native English speakers and learners of English showed the pattern. In the guess phrase task, native English speakers were able to generate 86% more correct guesses for sound symbolic words compared to non-sound symbolic words, and learners of English were able to generate 95% more correct guesses for sound symbolic words. In the recognition task, native English speakers were 25% more likely to recognize the meaning of sound symbolic words, and learners of English were 14% more likely to do so. Having these results, Parault & Parkinson (2008) concluded that both native English speakers and those learning English as a second language were able to recognize the definitions of sound symbolic words better than non-sound symbolic words.

However, there are some questions need to be considered. Each of the 13 clusters contains two to three different consonants that differ in phonetic features such as voicing and stop/continuant contrast as well as contrasting vowels. It needs to be investigated whether the effect found in this study is due to the sound symbolic clusters which are certain combinations of consonants or due to other consonants or vowels contained in the words.

Contrary to many of these studies in which cross-linguistic characteristics of sound symbolism were found, some aspects of sound symbolism are recognized to be language-specific. Iwasaki, Vinson, & Vigliocco (2007) found that there is disagreement in associating certain sounds with some evaluative dimensions. For instance, they found that Japanese native speakers tend to associate voiceless sounds with beauty, pleasantness or goodness, whereas English speakers with no experience in Japanese tend to associate voiced sounds with those dimensions.

Iwasaki, Vinson, & Vigliocco (2007) examined sound symbolism in Japanese sound-imitating words (*giongo*) and mimetic words (*gitaigo*). They investigated whether speakers of other languages are able to access the meaning of the two types of Japanese onomatopoeic words from their sounds. They asked native English speakers who had no prior knowledge of Japanese to listen to the target words and rate each word's meaning on semantic differential scales. One type of target word was sound-imitating words that express different manners of 'laughing'. 24 words were selected that consist of 17 CV-based words and 7 CVCV-based words. 13 of them referred to loud laughter such as *ahaha* and *wahaha,* whereas 11 of them referred to quiet suppressed laughter such as *uhuhu* and *huhu.* Another type of target word was action-imitating mimetic words that express different manners of 'walking'. There were 28 words that were largely CVCV-based because the most common mimetic words for walking are CVCV-based. All target words were rated along 20 semantically different 7-point scales for 'laughing' and 21 scales for 'walking' (e.g. big-small, good-bad, graceful-clumsy). Japanese speakers were given printed target words and English speakers listened to recorded target words pronounced by a native Japanese speaker.

The ratings for 'laughing' words by both Japanese and English native speakers were significantly correlated for 12 out of 24 of target words used in this experiment which included a mixture of both loud and quiet laughter. The most highly correlated ratings were for *kusu-kusu* 'quiet, restrained, non-resonant laughter' whose averaged ratings are organized shown below (Figure 12).

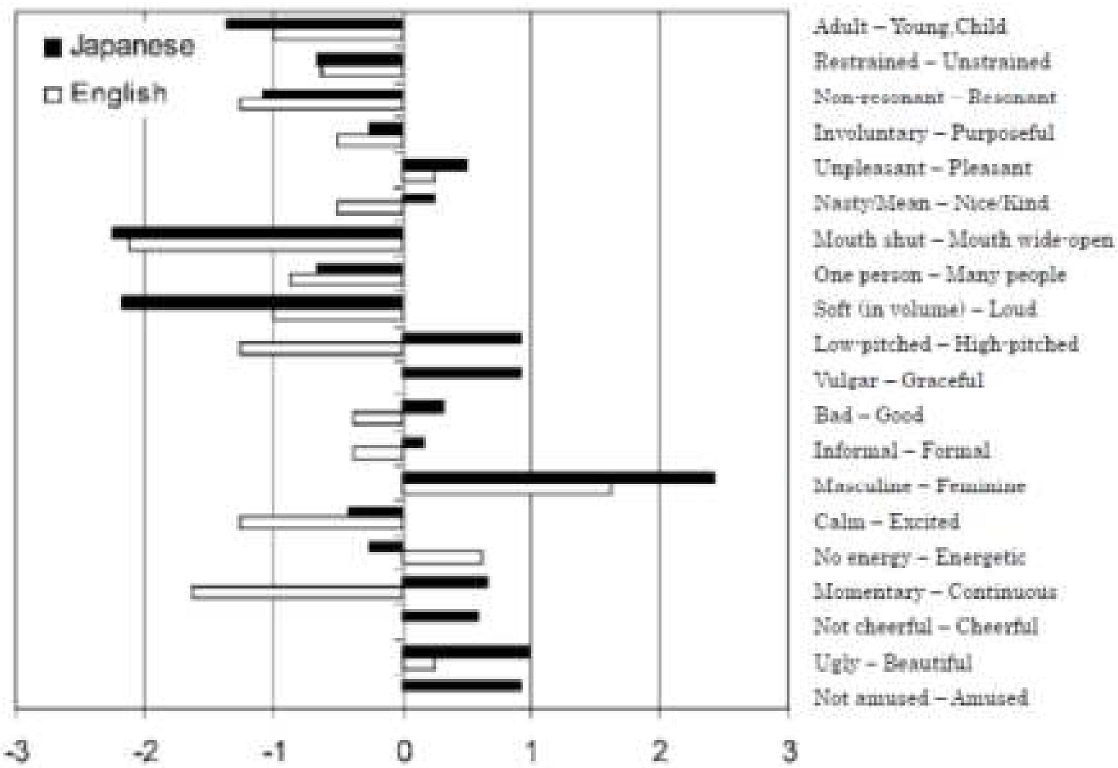

Figure 12: Average ratings by Japanese and English speakers for *kusu-kusu* from Iwasaki et al. (2007)

Overall, the most significantly correlated ratings were on the following dimensions; Loud (r=0.78, p<.001), followed by Mouth-Wide-Open (r=0.72, p<.001), Continuous (r=0.68, p<.001), and Resonant-Voice (r=0.63, p<.001).

The ratings by Japanese speakers varied considerably depending on the vowel and geminates. For example, Japanese speakers tended to associate /a/ with amused, cheerful, energetic, excited, loud, mouth-wide-open, and resonant-voice which were all found to be uncharacteristic of /u/. English speakers' ratings were similar to those of Japanese for the vowel /a/ and /u/. However, English speakers' ratings exhibited less definite differences among the vowels. As for geminate, Japanese speakers associated /Q/ with abrupt, short, forceful movement whereas English speakers showed no such effect. This was considered to be due to insensitivity of English speakers to Japanese phonological properties.

The ratings for 'walking' words by Japanese and English native speakers showed different results from that of 'laughing'. For the Japanese and English ratings, there were much fewer mimetic words for walking that were significantly correlated; only 7 out of 28 target words showed correlated rating patterns between Japanese and English speakers. The most highly correlated of these was *toko-toko* (r=0.69, p<.001) 'a small person walking quickly'. The averaged ratings are organized in the figure below (Figure 13).

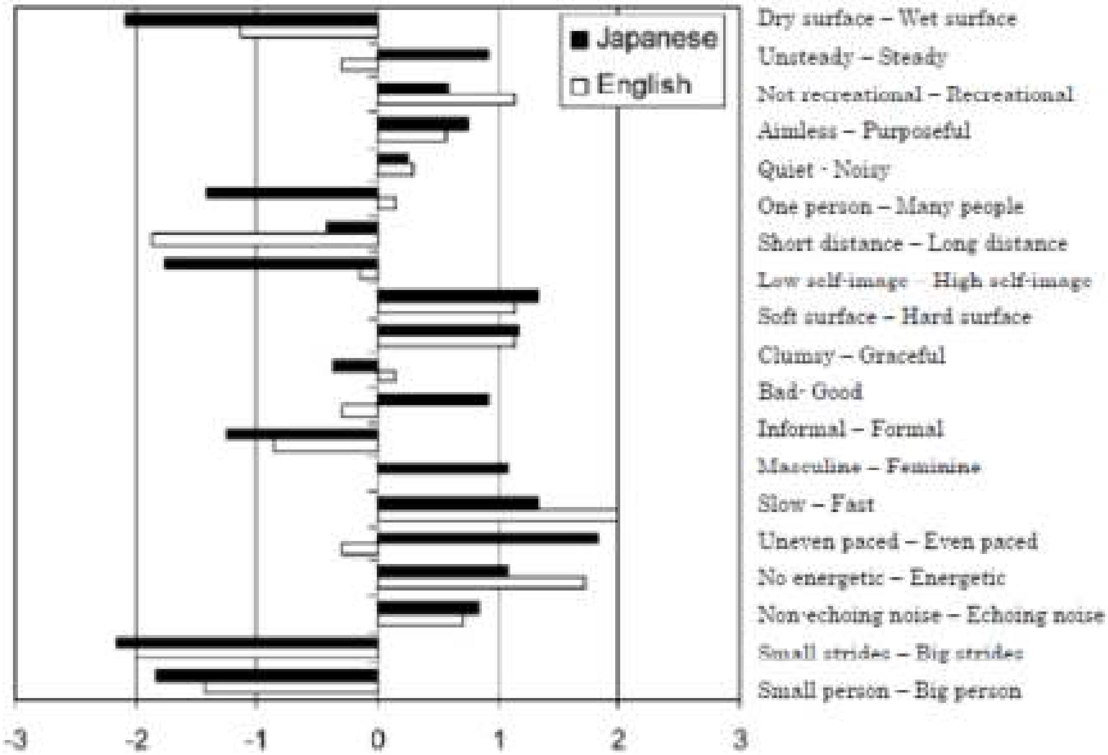

Figure 13: Average ratings by Japanese and English speakers for *toko-toko* from Iwasaki (2007)

Overall, Japanese and English speakers' ratings for 'walking' were correlated for only a small number of semantic dimensions which are Hard-Soles (r=0.37, p=.026) and Wet-Surface (r=0.32, p=.048).

There was also fewer vowel effects compared to 'laughing' words for both subject groups. Iwasaki et al. (2007) suggested that this indicates that the sound-meaning relationships in vowels are not as robust for primarily visual manner mimetic words such as 'walking' compared to primarily audio manner mimetic words such as 'laughing'.

Iwasaki et al. (2007) also compared the perception of voiced onset (words starting with a voiced consonant) and voiceless onset (words starting with a voiceless consonant). Both Japanese and English speakers showed a similar rating pattern only for the Big-Small dimension. They associated voiced onset with bigness and voiceless onset with smallness. In addition to Big-Small dimension, Japanese speakers showed a significant voicing effect on 7 other dimensions (see Figure 14). For instance, Japanese speakers tended to associate a voiceless onset with Good and Graceful, and a voiced onset with Bad and Clumsy. These effects were only significant with Japanese speakers.

Voiced onset	Voiceless onset
Big strides	Small strides
Uneven-paced	Even-paced
Masculine	Feminine
Informal	Formal
Bad	Good
Clumsy	Graceful
Low self-image	High self-image
Noisy	Quiet

Figure 14: Japanese perceptional tendency for words beginning with voiced and voiceless consonants from Iwasaki et al. (2007)

To sum up, English and Japanese speakers' responses to mimetic words correlated for a number of words and different semantic dimensions. The correlation is more significant for audio manner mimetic words such as 'laughing' words than for visual manner mimetic words such as 'walking' words. Iwasaki et al. (2007) suggested that this might be due to the fact that manners of laughing are more iconic and phonetically easier to express (e.g. the opening of the mouth) than manners of walking. They claimed that sound symbolism such as big-small, masculine-feminine, noisy-quiet, wet-dry, and soft-hard symbolism are shared across languages but some symbolism such as beautiful-ugly, graceful-clumsy, and good-bad are demonstrated only by Japanese speakers, which suggests that some sound symbolism is language-specific. Having the results that Japanese and English speakers responded oppositely for some evaluative dimensions such as beauty and pleasantness, they also claimed that language-specific sound symbolism exists on some evaluative dimensions since Japanese speakers tend to associate gracefulness and goodness with voiceless sounds whereas English speakers tend to associate those with voiced sounds.

2.4. Sound symbolism in Japanese mimetic words

Some word classes of particular languages have been found to be sound symbolic. One classic example is Japanese mimetic words (*gitaigo* and *giongo*). *Giongo* are onomatopoetic words that imitate sounds (e.g. “bowwow” in English) whereas *gitaigo* are words that imitate physical mode such as actions and psychological or physiological states (e.g. “boing-boing” in English). *Gitaigo* also refers to experiences that are related to vision, touch, taste, and olfaction. Japanese mimetic words are all semantically very specific and there are no hyponyms or hypernyms among them (Watson 2001). None refers to superordinate concepts such as walking and running in the general sense (Kita 2008). For instance, there are different verbs to specify certain manners of walking in

English such as “to wobble”, “to stroll”, and “to toddle”. The Japanese equivalents of these verbs are compositional as they are comprised of combinations of a generic verb（*aruku* ‘walk’） and mimetic words (*yoro-yoro* ‘unsteadily’, *bura-bura* ‘aimlessly’, and *yoti-yoti* ‘totteringly’) that are often used as adverbs to achieve a similar effect to those hyponyms (see examples 5, 6, 7 below). The following are examples of how mimetic words are used to describe different manners of walking.

5. ***yoro-yoro*** aruku. ‘to wobble’
gitaigo ‘walk’

6. ***bura-bura*** aruku. ‘to stroll’
gitaigo ‘walk’

7. ***yoti-yoti*** aruku. ‘to toddle’
gitaigo ‘walk’

The speech category of mimetic words in Japanese is difficult to identify because the categories depend on the linguistic environment in which it appears (Tsujimura 2007). They can be adverbs, adjectives, nouns, and adverbial verbs when attached to a light verb *suru* ‘do’. Mimetic adverbs are more numerous than adjectives and verbs. The major types of mimetic adverbs are those based on CV roots and those based on CVCV roots (example 8 and 9).

8. *pakatt* ‘the state of something’s mouth being open’

9. *paka-paka* ‘sound of hooves, the air going in and out’

As shown in example 9, mimetic words in Japanese often involve repetition of the first portion. 41% of mimetic words are in the repetition form [CV-CV] or [CVCV-CVCV] as in 9, 16% are [CV(CV)tt] as in 8, 9% are [CV(CV)ri], 8% are [CVttCVri], and 8% are [CV(CV)n]. Since the majority are [CV-CV] or [CVCV-CVCV] form, we will focus on this repetition form in the current paper.

Unlike Indo-European languages, which have few mimetic words, Japanese has the second largest number of mimetic words following Korean. It is said that there are over 5000 mimetic words in Japanese, and the number has been increasing (BMFT Publisher Japan 2012). Mimetic words are commonly used in daily lives among native speakers of Japanese since they provide speakers with rich means of expression that reveal subtle sensitivity. Their expressive meanings are immediately understood, and most expressions are readily identifiable for native speakers of Japanese (Hamano 1998). Therefore, mimetic words are frequently used in daily conversations and also used in a wide range of market outlets such as newspapers, comic books, novels, and magazines.

Japanese mimetic words have been attracting attention for the semantic impact that they provide. The number of mimetic words has been increasing and the number of people's use of mimetic words has been increasing in the past 20 years (NHK Close-up Gendai, June 2013). According to a database of the Japanese Diet Record, the usage count of mimetic words at the National Diet has increased to more than double in 2011 as compared to 1990 (Osaka University, 2011). This is because of the rich particular image a mimetic word carries in its sound. For example, an adjective 'soft' is *yawarakai* in Japanese. There used to be one mimetic word that expresses softness, *fuwa-fuwa*. However, another mimetic word, *mofu-mofu* has been created recently (Denki Tsushin University, Japan). Both *fuwa-fuwa* and *mofu-mofu* express something soft, but *mofu-mofu* expresses warmth and thickness that *fuwa-fuwa* does not express (Denki Tsushin University, Tokyo). Many food companies have started naming their products using mimetic words because people tend to buy products with sound that relates to their meaning. For instance, people prefer bread called "*Fuwa-fuwa* Bread" to a bread simply called "Bread" (BMFT "Words expressing tastes", 2012).

Hamano (1998) claimed that Japanese mimetic words constitute a highly cohesive system with distinct phonological, semantic, and syntactic characteristics. According to Hamano (1998), there are two grammatical categories of mimetic words that form sound-symbolic system; mimetic adverbs and mimetic nominal adjectives. The following figure shows examples of her findings on vocalic sound symbolism of mimetic adverbs based on CV_1CV_2 roots. She claimed that the second vowel (V_2) in CV_1CV_2 roots represents certain meaning.

	protrusion	Line/tenseness	small	large
/i/	-	+	-	-
/u/	+	-	+	-
/o/	-	-	-	-
/a/	-	-	-	+
/e/	vulgarity			

Figure 15: Vocalic sound symbolism in V_2 position from Hamano (1998)

Example 10 and 11 demonstrate size symbolism found in /u/ and /a/ at V_2 position. The vowel /u/ tends to express smallness whereas /a/ expresses largeness. The size difference can be described in the mimetic words that express certain manners of falling in the following sentences.

10. Ame-ga atama ni ***potu**-tto* otita.
    rain-NOM head on fell
    'A rain drop fell on my head and spread in a small patch.'

11. Ame-ga atama ni ***pota**-tto* otita.
    rain-NOM head on fell
    'A rain drop fell on my head and spread in a big patch.'

The example 12 shows that the vowel /i/ is associated with 'thinness/ slenderness/ sharpness/ tenseness'. It has /i/ at V_2 position expressing tenseness.

12. Kare-wa sugu ***puri-puri*** suru.
    He-NOM soon do
    'He gets irritable soon.'

(Hamano, 1998)

The vowel /e/ in V_2 position tends to show 'inappropriateness/ vulgarity'. The example 13 has /e/ in V_2 position expressing 'inappropriateness/ vulgarity'.

13. Nani-ga okottano-ka wakara-nakute ***poke****-tto* mitei-ta.
What-NOM happened-QUES understand-NEG watching-PAST
'Not realizing what had happened, he stood there with a stupid look.'
(Hamano, 1998)

Vowel length expresses the distance or duration that is involved in the movement.

14. ***Supon****-to* koruku-ga nuketa.
Cork-NOM came off
'The cork came off with a (light) pop.'

15. ***Supoon****-to* koruku-ga nuketa.
Cork-NOM came off
'The cork came off with a (forceful and long) pop.'
(Hamano, 1998)

(ii) Voicing contrast of obstruents

In C_1VC_2V-based mimetic adverbs, voicing contrast of obstruents is more remarkable on the intial consonant C_1 than the second consonant C_2 (Hamano, 1998). The voicing contrast on C_1 involves a semantic contrast in terms of 'mass/weight'. Voiceless obstruents such as /p, t, k, s/ symbolize lightness and smallness whereas voiced obstruents such as /b, d, g, z/ symbolize heaviness and largeness. Examples 16, 18, 20, and 22 have a voiceless obstruent in C_1 position whereas examples 17, 19, 21, and 23 have a voiced obstruent. The voicing contrast of example 16 and 17, 18 and 19, 20 and 21, and 22 and 23 affect the size and weight of the object which is the subject of each sentence.

16. Ame-ga ***para-para*** hutte kita.
rain-NOM fall came
'It started to sprinkle.'

17. Hyoo-ga ***bara-bara*** hutte kita.
hail-NOM fall came
'Hailstones began falling in big heavy drops.'

18. ***Toro**-tto* sita hatimitu
did honey
'thick and smooth honey'

19. ***Doro**-tto* sita mizu
did water
'murky water'

20. ***Kuru-kuru*** mawaru dai arigatoo.
spin tray thanks
'Thank you for the lightly-spinning tray'

21. Kaitendoa-ga ***guru-guru*** mawatte naka ni haire-nakat-ta.
revolving door-NOM rotate inside to enter-NEG-PAST
'The revolving door heavily turned around and around, and I couldn't get in.'

22. ***sara-sara*** sita sio
did salt
'finely processed salt'

23. ***zara-zara*** sita sio
did salt
'coarse salt'

The voicing contrast in mimetic words extends to psychological conditions. The voiceless obstruents can have the meaning of 'light feelings' (examples 24 and 26) whereas the voiced obstruent are often associated with 'unpleasantness' and 'heavy feelings/ suspicion/ complaints' (examples 25 and 27).

24. ***Sito-sito*** ame-ga huru.
Rain-NOM fall
'Soft rain falls.'

25. ***Zito-zito*** ase-ga detekuru.
sweat-NOM come out
'I get sticky with sweat.'

26. ***Kara-tto*** sita hito
did person
'a jolly and open-minded person'

27. ***Boso-tto*** sita hito
did person
'a cheerless person'

(iii) Palatalization

Palatalization involves a semantic continuum of 'childishness/immaturity/unreliability' and 'diversity/excessive energy/lack of restraint'. The following is a minimal pair that shows the contrast in palatalization.

28. *pata-pata* 'flapping'

29. *patya-patya* 'splashing water' (involves 'excessive energy')

Research on sound symbolism in Japanese has been explored extensively by Hamano (1998), but mimetic words have recently started to attract a lot of attention of researchers examining natural language processing (Komiya & Kotani 2008). Not all the findings by Hamano (1998) however are in agreement with recent studies on sound symbolism in Japanese mimetic words such as Ivanova (2006).

Ivanova (2006) examined sound symbolism in Japanese mimetic words that are produced in native Japanese speech or appear in acknowledges Japanese mimetic word dictionaries. She proposed 37 phonaesthematic patterns typical for the majority of the Japanese mimetic words. She claimed that sound-symbolic effect in Japanese mimetic words is produced by two major factors: sequences of specific phonemes and their particular positions in words. According to the two factors, 37 phonaesthematic patterns were grouped into 5 types: (i) The first (and third) mora, (ii) The vowel of the first mora and the whole second mora, (iii) The second and fourth moras, (iv) All the moras but the first, (v) Partial reduplication.

Type (i) consists of 6 patterns of words sharing the first mora or its reduplication in the third mora. The example (i) is one of the 6 patterns.

(i) [**mu** + CV + **mu** + CV] or [**mu** + X] “excessive energy”, “suppression”
As in: ***muka-muka*** ‘retch’, ***mun-mun*** ‘stuffy, sultry’

Type (ii) consists of 7 patterns of words sharing the first mora and the whole second mora with completely reduplicated CVCV-roots. The example (ii) is one of the 7 patterns.

(ii) [C + **asa** + C + **asa**] “disappointing appearance due to some definiency”
As in: ***basa-basa*** ‘crumbly’, ***kasa-kasa*** ‘dry, rough’

Type (iii) consists of 12 patterns of words sharing the second and fourth moras. The example (iii) is one of the 12 patterns.

(iii) [CV + **bu** + CV + **bu**] “a great amount of liquid or flesh swaying”
As in: *de**bu**-de**bu*** ‘fat and flabby’, *ga**bu**-ga**bu*** ‘guzzle/chug’

Type (iv) consists of 10 patterns of words with no reduplication. Their typical structure are [C_1V_1 + /Q/ + C_2V_2 +ri] and [C_1V_1 + /N/ + C_2V_2 + ri].

(iv) [CV + **N + wa + ri**] “gentleness”
As in: *fu**nwari*** ‘fluffy’, *ji**nwari*** ‘well up slowly’

Type (v) consists of 2 patterns of words with different first and third moras, and second mora reduplicated in the fourth.

(v) [C_1V_1 + **cha** + C_2V_2 + **cha**] “disorder, unattractiveness”
As in: *me**cha**-ku**cha*** ‘unreasonable, incoherent’, *pe**cha**-ku**cha*** ‘chatter’

Ivanova (2006)’s approach to sound symbolism in Japanese mimetic words is different from that of Hamano (1998)'s in a sense that Ivanova (2006) found sound symbolism in a certain mora (sequence of phonemes) or a combination of moras, whereas Hamano (1998) found sound symbolism in phonetic features such as voicing in obstruents and height/frontness of vowels, not in combinations of those. Therefore, some findings by Ivanova (2006) overlaps Hamano (1998)'s, but not all the 37 phonaesthematic patterns match Hamano's observations. For instance, Hamano (1998) claimed the consonant /m/ represents 'murkiness' and the vowel /u/ represents 'smallness'. Ivanova (2006) introduced a pattern belonging to Type (i) which contains a mora /mo/ (a combination of /m/ and /o/). Ivanova (2006) found that the mora /mo/ represents 'murkiness'. This observation matches Hamano (1998)'s. However, Ivanova (2006) introduced another pattern of Type (i) which contains a mora /mu/ (a combination of /m/ and /u/) that represents 'excessive energy'. This does not match Hamano (1998)'s finding because the vowel /u/ represents 'smallness' and not 'excessiveness' as in Hamano (1998).

Japanese mimetic words constantly are one of word classes that are found to be sound symbolic because of the specific semantic information and iconic images they carry and their frequent use by Japanese speakers. According to studies on Japanese mimetic words such as Hamano (1998) and Ivanova (2006), particular phonemes are systematically used to express certain meanings. There is some agreement in their findings, but not all are congruent. Some of Hamano (1998)'s findings such as voicing contrast in consonants are found to be a sound symbolic in other languages such as English and Chinese (Shinohara & Kawahara, to appear). The current project chose to investigate sound symbolism in the voicing contrast because it is a distinction in both Japanese and in English (e.g. voiced=big, voiceless=small) that have been studied.

3. Purpose of the current study

The current study investigated sound symbolism in Japanese using Japanese mimetic non-words. In this study, voicing of consonants was examined and vowel influence was controlled. We examined whether the voicing contrast in consonants (/t, k, s/ vs. /d, g, z/) affects the perception in both Japanese native speakers and English native speakers who had no knowledge of Japanese. Two additional manipulations were also included. First, stimuli were evaluated on 4 different dimensions including both size and shape dimensions as well as evaluative dimensions, in order to examine the generality and shares of the sound symbolism. Second, voicing was manipulated, creating a continuum from voiced to voiceless endpoints, in order to examine the categorical nature of the perception.

The stimuli were 3 pairs of novel Japanese mimetic non-words that contrast in the voicing of the first and third consonants (*deze deze & tese tese, gede gede & kete kete, zege zege & seke seke*). Unlike Iwasaki et al. (2007) in which they used real mimetic words as stimuli, we used mimetic non-words in order to avoid associations of the real words. By using non-words, we are able to test whether sound symbolism exists in certain sounds (e.g. a voiceless consonant /k/) but not in existing words (e.g. *kusu kusu 'to giggle'*). In order to put focus on the voicing contrast of consonants, we controlled vowel contrast by using only one vowel /e/ in the stimuli. This was due to pervious research that showed that the height and the frontness of vowels influence how the sound is perceived (Hamano (1998), Wremberl (2010), and Shinohara & Kawahara (to appear)). The vowel /e/ is the only vowel among Japanese vowels (/i, a, e, o, u/) that is neither extremely high or low nor front or back.

For each of the 3 pairs of mimetic non-words (*deze deze & tese tese, gede gede & kete kete, zege zege & seke seke*), a voicing continuum of 4 members was created. Each pair had a voiced member (*deze deze*), an ambiguous voiced member, an ambiguous voiceless member, and a voiceless member (*tese tese*). The ambiguous voiced member and the ambiguous voiceless member were created by manipulating the stop and fricative duration of the voiced and the voiceless member, respectively. The ambiguous voiced member was created from the voiced endpoint stimulus, making it more perceptually ambiguous in terms of its voicing. The ambiguous voiceless member was created from the voiceless endpoint stimulus, making it more perceptually ambiguous in terms of its voicing. Therefore, there were 3 pairs x 4 members = 12 stimuli in total. These voicing continua were created in order to investigate the nature of transition from voiced sounds to voiceless sounds.

For the current study, both Japanese and English participants were used. We compared the perception of Japanese speakers and English speakers on 4 different dimensions: big-small, round-spiky, good-bad, and graceful-clumsy. The big-small and round-spiky dimensions were tested in order to examine size symbolism and shape symbolism, respectively. These two types of major sound symbolism have been found to be cross-linguistic.

According to Hamano (1998), Shinohara & Kawahara (to appear), and Iwasaki et al. (2007), the voicing of consonants affects the perception of size; Japanese speakers tend to associate voiced sounds with largeness and voiceless sounds with smallness. This tendency is also found among speakers of other language such as English and Chinese speakers (Shinohara & Kawahara, to appear). Therefore, we expected to find similar perceptional patterns between Japanese and English speakers for the big-small dimension in that both language groups would associate voiced sounds with largeness and voiceless sounds with smallness.

As we introduced earlier, the stop vs. continuant contrast is found to affect the perception of shapes. Kohler (1929) and Westbury (2005) found that English speakers tend to associate continuant consonants with roundness and stop consonants with spikiness. Aveyard (2012) found that non-plosive consonants tend to represent roundness and plosive consonants tend to represent spikiness. The current study further examined whether the voicing of consonants affects the perception of shape among Japanese speakers and English speakers. We expected for both Japanese and English speakers to associate voiced sounds with roundness and voiceless sounds with spikiness. In addition to the voicing contrast, we also examined whether having stops or fricatives would affect perceived shape of an image. Previous studies found perceptional distinction between stops and continuants on the round-spiky dimension in English speakers, thus we expected English participants to associate stops with spikiness. There were no previous studies that examined shape symbolism in fricatives, therefore the current study explored whether fricatives would play any role in perceived shape of an image.

The other 2 dimensions, good-bad and graceful-clumsy, were chosen to test the finding by Iwasaki et al. (2007) that sound symbolism in evaluative dimensions tend to be language-specific. Both good-bad and graceful-clumsy were found to be language-specific in Iwasaki et al. (2007); Japanese speakers tended to associate voiced sounds with badness and clumsiness whereas English speakers tended to do so with voiceless sounds. The current study examined whether the voicing affects evaluative perception, and whether Japanese and English speakers share the perceptional tendency towards goodness and gracefulness. We expected for Japanese speakers to associate voiced sounds with badness and clumsiness, and voiceless sounds with goodness and gracefulness. Contrary, we expected for English speakers to associate voiced sounds with goodness and gracefulness, and voiceless sounds with badness and clumsiness.

The current study investigated three research questions. First, does the voicing contrast in consonants affect the perception of Japanese and English speakers? If so, on what semantic or evaluative dimensions does the voicing influence each language group? Second, does stop (/d, g, t, k/) and fricative (/z, s/) affect perceived shape of an image? Third, does manipulation of voicing continuum from voiced to voiceless affect perception on different dimensions? By conducting the current perception experiment, we investigated how speakers of different languages are able to access the same meaning from their sounds.

4. Methods

4.1. Participants

Participants were 12 native speakers of Japanese and 16 native speakers of English who had no prior knowledge of Japanese.

The Japanese speakers (N=12, 7 females, 5 males, average age: 33) were either recruited either on Lawrence campus at University of Kansas or in Fukuoka, Japan. 3 of them had lived in the U.S.A. for 2 years, 5 of them had lived in the U.S.A. for 6 months, and 4 of them had never lived outside Japan at the time of the experiment. Including the

5 participants who had never lived outside Japan, all 12 participants learned English at junior high school and high school. They volunteered to participate in the current experiment; no reward was given. All of the participants had no known hearing disorder. The participants were asked to read and sign a consent form for participating in a linguistic perception experiment beforehand.

All the native speakers of English (N=16, 12 females, 4 males, average age: 24) were recruited on Lawrence campus at University of Kansas. They were either college students or graduate students. 3 of them were English monolinguals, 12 of them were either beginner or intermediate learners of European languages such as Spanish, French, and Italian. One participant was fluent in Amharic. None had any knowledge of the Japanese language. The majority of them received course credit for their participation. All of the participants had no known hearing disorder. The participants were asked to read and sign a consent form for participating in a linguistic perception experiment beforehand.

4.2. Recordings

The novel mimetic words were recorded on a solid-state recorder (Marantz PMD671) using a cardioid microphone (Electrovoice-N/D-767) for noise-free recording in an anechoic chamber (IAC) on Lawrence campus at University of Kansas. The stimuli were pronounced by a female native speaker of Japanese who spoke the Tokyo dialect.

4.3. Materials

The stimuli consisted of 3 voiced-voiceless pairs that are novel Japanese mimetic non-words in C_1VC_2V-C_3VC_4V form (*deze deze & tese tese; gede gede & kete kete; zege zege & seke seke*). In order to maximize the voicing effect on sound symbolism, only one vowel /e/ was used in the stimulus words. According to Hamano (1998), Wrembel (2010), and Shinohara & Kawahara (to appear), the height and the frontness of Japanese vowels influence how the sound is perceived. Among the Japanese vowels /i, a, e, o, u/, only the vowel /e/ is neither extremely high or low nor front or back.

For the voiceless stimuli, three of the novel mimetic non-words consisted of one of voiceless consonants /t, k, s/ in C_1 and C_3 position and a different voiceless consonant (/t, k, s/) in C_2 and C_4 position (*tese tese, kete kete, seke seke*). For the voiced stimuli, the three of novel mimetic non-words consisted of one of voiced consonants /d, g, z/ in C_1 and C_3 position and a different voiced consonant (/d, g, z/) in C_2 and C_4 position (*deze deze, gede gede, zege zege*). It has been confirmed that none of those words exist in a Japanese dictionary of onomatopoeias ("Usage guide to Japanese onomatopoeias" by Atouda & Hoshino, 2009).

Among the three voiced-voiceless pairs, the *"gede-kete"* pair consisted of stops (/g, d, k, t/) and the vowel /e/, whereas the other two pairs, *"deze-tese"* and *"zege-seke"* consisted of stops, fricatives (/z, s/) and the vowel /e/.

For each of the 3 voiced-voiceless stimulus pairs (*deze deze & tese tese, gede gede & kete kete, zege zege & seke seke*), an ambiguous continuum was created for each pair by adjusting the stop and fricative duration of the first and the third consonants (e.g. /d/ in *deze deze*, /t/ in *tese tese*, /g/ in *gede gede,* /k/ in *kete kete,* /z/ in *zege zege*, /s/ in *seke seke*). This is to investigate categorical perception of voiced sounds to voiceless sounds. One of the two ambiguous members is perceptually closer to the voiced stimulus (ambiguous voiced) and the other is closer to the voiceless stimulus (ambiguous voiceless). Therefore, a 4 member continuum for each of the 3 voiced-voiceless pairs was tested.

The stop and fricative duration of the 3 recorded stimulus pairs (pronounced by a female Japanese speaker) were adjusted, creating the ambiguous continuum members using Praat (Boersma & Weenink, 2007: doing phonetics by computer, version 4.6.38) All the stimulus words had the same accentuation pattern of LHHH (low-high-high-high) which is the typical pattern for mimetic adjectives in CVCV-CVCV form. The recorded stimuli were all segmented to make 2 CVCV strings; the first part of the CVCV string has an accentuation of LH and the second repeated part of the string has HH (e.g. for *deze deze,* $de^{L}ze^{H}$ $de^{H}ze^{H}$).

4.3.1. Stop/fricative duration manipulations

4.3.1.1. *deze deze & tese tese* (/d/ and /t/ were manipulated):

The stop duration of the voiced and voiceless stimuli was measured. For the voiced stimulus, the prevoicing and burst release of the C_1 in the C_1VC_2V string "*deze 1"* was 62 ms. The closure and release of the C_3 in the C_3VC_4V string "*deze 2"* was 48 ms.

In order to make an ambiguous voiced member, the prevoicing and burst of the first string "*deze 1"* (62 ms) was removed, and then the burst and aspiration of the first string "*tese 1"* (40 ms) was appended. The appended segment was then shortened to 28ms by reducing the aspiration. The same procedure was done for the second repeated "*deze 2"* string. The closure and release of *"deze 2"* (48 ms) was removed, and then the closure and release of *"tese 2"* (65 ms) was appended. Stimulus segmentation was determined from both amplitude measures in the waveforms and perceptual evaluation of the stimuli. As a result of the manipulation, the stop duration of the first string of the ambiguous voiced "*deze"* became +28 ms, and the second repeated string of the ambiguous voiced "*deze*" resulted in 65 ms.

For the voiceless stimulus, the burst and aspiration of the C_1 in the C_1VC_2V string *"tese 1*" was +40 ms and the closure and release of C_3 of the C_3VC_4V string *"tese 2"* was +65 ms.

In order to make an ambiguous voiceless member, the aspiration of the C_1 of the

C_1VC_2V string *"tese 1"* and the closure of the C_3 of the C_3VC_4V string *"tese 2"* were both shortened to 23 ms and 40 ms respectively. Stimulus segmentation was determined from both amplitude measures in the waveforms and perceptual evaluation of the stimuli. As a result of the manipulation, the stop duration of the first string of the ambiguous voiceless "*tese 1"* became 23 ms and the second repeated string of the ambiguous voiceless *"tese 2"* became 40 ms.

In this way, a four member voicing continuum was created for both the initial C1 (62 ms, +28 ms, +23 ms, +40 ms) and the medial C3 (48 ms, +65 ms. +40 ms, +65 ms), beginning with an unmanipualted voiced endpoint stimulus, an ambiguous voiced stimulus, an ambiguous voiceless stimulus, and an unmanipulated voiceless endpoint stimulus.

4.3.1.2. *gede-gede & kete-kete* (/g/ and /k/ were manipulated):

The stop duration of the voiced and voiceless stimuli was measured. For the voiced stimulus, the prevoicing and burst release of the C_1 in the C_1VC_2V string "*gede 1"* was 90 ms and the closure and release of the C_3 in the C_3VC_4V string "*gede 2"* was 38 ms.

In order to make an ambiguous voiced member, the prevoicing and burst release of the first string "*gede 1"* (90 ms) was removed, and then the burst and aspiration of the first string "*kete 1"* (64 ms) was appended. The appended segment was then shortened to

38 ms by reducing the aspiration. The same procedure was done for the second repeated *"gede 2"* string. The closure and release of *"gede 2"* (38 ms) was removed, and then the closure and release of *"kete 2"* (94 ms) was appended. The appended closure was then shortened to 34 ms by reducing the closure. Stimulus segmentation was determined from both amplitude measures in the waveforms and perceptual evaluation of the stimuli. As a result of the manipulation, the stop duration of the first string of the ambiguous voiced "*gede 1"* became 38 ms, and the second repeated string of the ambiguous voiced "*gede 2*" became 34 ms.

For the voiceless stimulus, the burst and aspiration of the C_1 in the C_1VC_2V string *"kete 1*" was 64 ms and the closure and release C_3 of the C_3VC_4V string *"kete 2"* was 94 ms.

In order to make an ambiguous voiceless member, the burst and aspiration of the C_1 of the C_1VC_2V string *"kete 1"* (64 ms) and the closure and release of the C_3 of the C_3VC_4V string *"kete 2"* (94 ms) were both shortened to 43 ms and 63 ms by reducing the aspiration and the closure respectively. Stimulus segmentation was determined from both amplitude measures in the waveforms and perceptual evaluation of the stimuli. As a result of the manipulation, the stop duration of the first string of the ambiguous voiceless "*kete 1"* became 43 ms and the second repeated string of the ambiguous voiceless *"kete 2"* became 63 ms.

In this way, a four member voicing continuum was created for both the initial C1 (90 ms, 38 ms, 43 ms, 64 ms) and the medial C3 (38 ms, 34 ms. 63 ms, 94 ms), beginning with an unmanipualted voiced endpoint stimulus, an ambiguous voiced stimulus, an ambiguous voiceless stimulus, and an unmanipulated voiceless endpoint stimulus.

4.3.1.3. *zege-zege & seke-seke* (/z/ and /s/ were manipulated):

The fricative duration of the voiced and voiceless stimuli was measured. For the voiced stimulus, the fricative duration of the C_1 in the C_1VC_2V string "*zege 1"* was +103 ms and the C_3 in the C_3VC_4V string "*zege 2"* was 96 ms.

In order to make an ambiguous voiced member, the fricative duration of the first string "*zege 1"* was removed, and then the fricative duration of the first string "*seke 1"* (150 ms) was appended. The appended segment was then shortened to 75 ms by removing the first half of the segment. The same procedure was done for the second repeated *"zege 2"* string. The fricative duration of *"zege 2"* (96 ms) was replaced by the fricative duration of *"seke 2"* (96 ms) that was shortened to 60 ms by removing the initial part of the segment. Stimulus segmentation was determined from both amplitude measures in the waveforms and perceptual evaluation of the stimuli. As a result of the manipulation, the fricative duration of the first string of the ambiguous voiced "*zege 1"* became 75 ms, and the second repeated string of the ambiguous voiced "*zege 2*' became 60 ms.

For the voiceless stimulus, the fricative duration of the C_1 in the C_1VC_2V string *"seke 1"* was 150 ms and the C_3 of the C_3VC_4V string *"seke 2"* was 96 ms.

In order to make an ambiguous voiceless member, the fricative duration of the C_1 of the C_1VC_2V string *"seke 1"* (150 ms) and the C_3 of the C_3VC_4V string *"seke 2"* (96 ms) were both shortened to 70 ms and 53 ms respectively by removing the initial segments of each fricative duration. Stimulus segmentation was determined from both amplitude measures in the waveforms and perceptual evaluation of the stimuli. As a result of the manipulation, the fricative duration of the first string of the ambiguous voiceless "*seke 1"* became 70 ms and the second repeated string of the ambiguous voiceless *"seke 2"* resulted in 53 ms.

In this way, a four member voicing continuum was created for both the initial C1 (103 ms, 75 ms, 70 ms, +150 ms) and the medial C3 (96 ms, 60 ms. 53 ms, 96 ms), beginning with an unmanipualted voiced endpoint stimulus, an ambiguous voiced stimulus, an ambiguous voiceless stimulus, and an unmanipulated voiceless endpoint stimulus.

Figure 16: Stop/fricative duration of the first and third consonants (same consonant) for each continuum

Stop / fricative duration of the C_1 of the first string (ms)

	voiced	Ambiguous voiced	Ambiguous voiceless	voiceless	
deze	62	28	23	40	**tese**
gede	90	38	43	64	**kete**
zege	103	75	70	150	**seke**

Stop / fricative duration of the C_3 of the second string (ms)

	voiced	Ambiguous voiced	Ambiguous voiceless	voiceless	
deze	48	65	40	65	**tese**
gede	38	34	63	94	**kete**
zege	96	60	53	96	**seke**

4.4. Procedure

The participants were first asked to answer a questionnaire on their language background. It was confirmed that none of the English speakers had any prior knowledge of Japanese. Then the participants were instructed to sit in front of a computer screen, wear headphones, and were provided with 4 answer sheets (one for each dimension). Instructions were first shown on the screen which was followed by two practice trials.

In the instructions, Japanese speakers were told that they would listen to some Japanese non-words and English speakers were told that they would hear some words of an unknown language. Both participants were directed to listen to a word by clicking a speaker icon shown on the screen and they were allowed to listen to the stimuli as many times as they wanted to. The instructions were in Japanese for Japanese speakers and in English for English speakers. After the practice trials, the target stimuli were played through the headphone as the participant clicked on the speaker icon.

The participants listened to each word and then circled one of the four points on the scale (big, relatively big, relatively small, small), (round, relatively round, relatively spiky, spiky), (good, relatively good, relatively bad, bad), (graceful, relatively graceful, relatively clumsy, clumsy). There were 3 voiced-voiceless pairs of non-words and each pair contained 4 continuum members. Those 12 (3 pairs x 4 continuum members) stimuli were presented randomly in each of the 4 dimensions. The participants were allowed to take a break. The whole procedure took about 30 minutes for Japanese speakers and 45 minutes for English speakers to complete.

Each novel mimetic word was rated along 4 dimensions; big-small dimension, round-spiky dimension, good-bad dimension, and graceful-clumsy dimension. For each dimension, a four-point scale was created and each point was indicated with the degree as in "big", "relatively big", "relatively small", and "small" for the big-small dimension, "round", "relatively round", "relatively spiky", and "spiky" for the round-spiky dimension, "good", "relatively good", "relatively bad", and "bad" for the good-bad dimension, and "graceful", "relatively graceful", "relatively clumsy", and "clumsy" for the graceful-clumsy dimension. The scales were later converted to 1 to 4 to express polarity with positive values corresponding to the label of the scale (1: big, 2: relatively big, 3: relatively small, 4: small), (1: round, 2: relatively round, 3: relatively spiky, 4: spiky), (1: good, 2: relatively good, 3: relatively bad, 4: bad), (1: graceful, 2: relatively graceful, 3: relatively clumsy, 4: clumsy). 12 four-point scales were written on a sheet of paper (answer sheet) for the participants to rate each of the twelve novel mimetic stimuli.

Microsoft Power Point was used to play the stimuli. One stimulus word was embedded to a Power Point slide; therefore there were 12 (3 pairs x 4 continuum members) PPT slides for each dimension, 48 (12 stimuli x 4 dimensions) pages per session. In order to gain reliability, there were two successive repeated sessions per participant. The order of the stimuli in both sessions was randomized for each dimension and for each participant.

The rating results for all 3 voiced-voiceless stimulus pairs on four-point scales were converted to 1 to 4 to express polarity with positive values corresponding to the label of the scale (e.g. 1: big, 2: relatively big, 3: relatively small, 4: small). The converted numbers were averaged across participants within each subject group. In each subject group, there were averaged voiced results, ambiguous voiced results, ambiguous voiceless results, and voiceless results for each of the 4 dimensions.

5. Results

A repeated measures 2 x 4 ANOVA (language x voicing) was conducted for each dimension (big-small, round-spiky, good-bad, graceful-clumsy). Language (Japanese and English) was a between subjects factor and voicing (voiced, ambiguous voiced, ambiguous voiceless, voiceless) was a within subjects factor.

5.1. Big-small dimension

For the big-small categorization task, participants judged whether the stimulus was big or small on a 4-point scale. Numeric values from 1 to 4 were assigned to each member of the scale, with 1=big, 2=relatively big, 3=relatively small, and 4=small.

The overall mean rating was 2.43 for the 12 Japanese participants. The overall mean rating was 2.40 for the 16 English participants. There was no significant difference across language background for the big-small dimension; $F\ (1, 26) = 0.17$, $p = 0.684$. The ratings for the Japanese and the English participants are shown in figure 17 for the big-small categorization task.

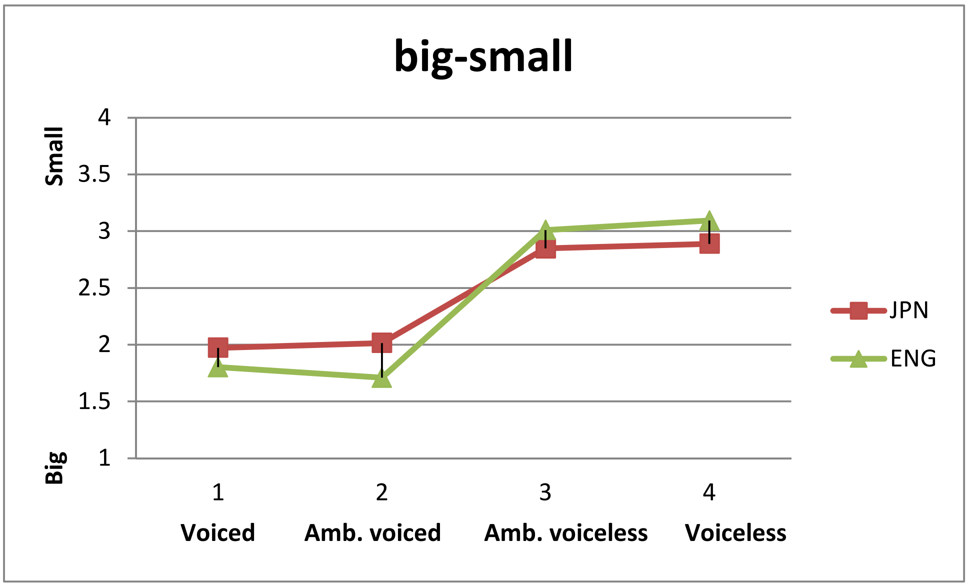

Figure 17: The mean rating (1-4) for Japanese and English speakers for the big-small dimension

The mean rating for the 12 Japanese speakers for the voiced stimuli (*deze deze, gede gede, zege zege*) was 1.97, for the ambiguous voiced stimuli 2.01, for the ambiguous voiceless 2.85, and for the voiceless stimuli (*tese tese, kete kete, seke seke*) 2.89.

The mean rating for the 16 English speakers for the voiced stimuli (*deze deze, gede gede, zege zege*) was 1.80, for the ambiguous voiced stimuli 1.71, for the ambiguous voiceless stimuli 3.01, and for the voiceless stimuli (*tese tese, kete kete, seke seke*) 3.09.

There was a significant main effect of voicing; $F(3, 78) = 57.79$, $p<0.001$. Planned comparison tests revealed that there was a significant voicing effects between the voiced and voiceless endpoint stimuli ($p<0.001$). In addition, significant differences were observed between the voiced and the ambiguous voiceless stimuli ($p<0.001$), and between the voiceless and the ambiguous voiced stimuli ($p<0.001$), as well as between the ambiguous voiced and the ambiguous voiceless stimuli ($p<0.001$). There was no significant voicing difference between the voiced and the ambiguous voiced stimuli ($p=0.78$) and between the ambiguous voiceless and voiceless stimuli ($p=0.51$).

While there was not an overall significant voicing x language interaction; $F(3, 78) = 2.30$, $p=0.084$, the English participants did show a slightly greater effect of voicing ($p=0.095$) than the Japanese participants which is due to a greater voicing x language effect for English speakers between the ambiguous voiced and the ambiguous voiceless stimuli. The English participants showed a difference of 1.30 (English ambiguous voiced 3.01 versus English ambiguous voiceless 1.71) as compared to only a .84 difference for the Japanese participants (Japanese ambiguous voiced 2.85 versus Japanese ambiguous voiceless 2.01).

Overall, for the big-small dimension, both Japanese speakers and English speakers showed a significant voicing effect and there was neither a significant language effect nor a language x voicing interaction. These results suggest that both Japanese speakers and English speakers associated both the voiced (*deze deze, gede gede, zege zege*) and the ambiguous voiced stimuli with bigness, and they associated the voiceless (*tese tese, kete kete, seke seke*) and ambiguous voiceless stimuli with smallness.

5.2. Round-spiky dimension

For the round-spiky categorization task, participants judged whether the stimulus was round or spiky on a 4-point scale. Numeric values from 1 to 4 were assigned to each member of the scale, with 1=round, 2=relatively round, 3=relatively spiky, and 4=spiky. Among the three stimulus pairs, the *"gede-kete"* pair was consisted of only stops /g, d, k, t/ and the vowel /e/, and the other two pairs *"deze-tese"* and *"zege-seke"* consisted of a stop, a fricative, and the vowel /e/.

The overall mean rating was 2.87 for the 12 Japanese participants. The overall mean rating was 2.45 for the 16 English participants. There was a significant difference across language background for round-spiky dimension; $F(1, 26) = 24.11$, $p<0.001$, with the Japanese participants having overall more spiky responses than the English participants. The ratings for the Japanese and the English participants are shown in figure 18 for the round-spiky categorization task.

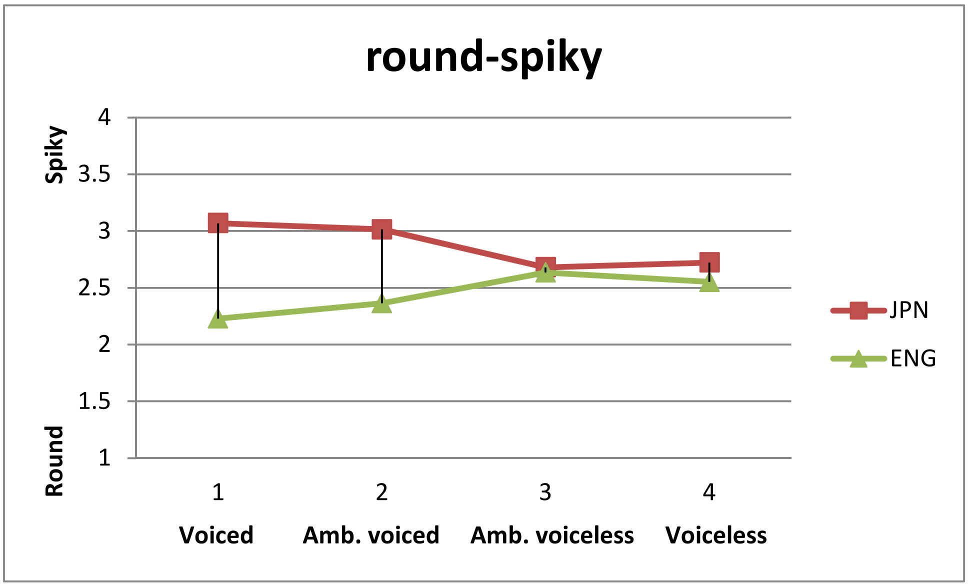

Figure 18: The mean rating for each stimulus for Japanese and English speakers

The mean rating for the 12 Japanese speakers for the voiced stimuli (*deze deze, gede gede, zege zege*) was 3.07, for the ambiguous voiced stimuli 3.00, for the ambiguous voiceless 2.68, and for the voiceless stimuli (*tese tese, kete kete, seke seke*) 2.72.

The mean rating for the 16 English speakers for the voiced stimuli (*deze deze, gede gede, zege zege*) was 2.23, for the ambiguous voiced stimuli 2.37, for the ambiguous voiceless stimuli 2.64, and for the voiceless stimuli (*tese tese, kete kete, seke seke*) 2.55.

Overall, there was no significant main effect of voicing; $F(3, 78)= 0.024$, $p=0.995$. No significant difference in voicing was observed between the voiced and voiceless endpoint stimuli ($p=0.951$). No significant difference was observed between the voiced and the ambiguous voiceless stimuli ($p=0.963$), and between the voiceless and the ambiguous voiced stimuli (p=0.838), as well as between the ambiguous voiced and the ambiguous voiceless stimuli (p=0.910). There was no significant voicing difference between the voiced and the ambiguous voiced stimuli (p=0.711) and between the ambiguous voiceless and voiceless stimuli (p=0.790).

However, only the English participants showed a significant main effect of voicing in the only-stop stimulus *"gede-kete"* pair. The English participants associated voiced stops (/g, d/) with roundness and voiceless stops (/k, t/) with spikiness. There was a significant difference observed between the two ambiguous stimuli (p=0.076) as well as between the two endpoints (p=0.001). There was no significant voicing difference between the voiced and ambiguous voiced (p=0.256) and between the voiceless and ambiguous voiceless (p=0.638). No such voicing effect was observed either in the two mixed (stop & fricative) pairs *"deze-tese"* or *"zege-seke"*. The Japanese participants did not show the voicing effect in any of the three stimulus pairs.

There was also no overall significant voicing x language interaction; F (3, 78)= 2.36, p=0.078.

To sum up, for the round-spiky dimension, the Japanese speakers did not show a significant voicing effect in any of the three stimulus pairs, whereas the English speakers showed a significant voicing effect only in the only-stop stimulus pair *"gede-kete"*. In the current study, the English participants associated voiced stops with roundness and voiceless stops with spikiness. No such tendencies were observed when there are both stops and fricatives in a stimulus. In general, the Japanese participants used more spiky responses. While the Japanese participants showed a slightly different pattern across the voicing continua, none of the differences was significant.

5.3. Good-bad dimension

For the good-bad categorization task, participants judged whether the stimulus was good or bad on a 4-point scale. Numeric values from 1 to 4 were assigned to each member of the scale, with 1=good, 2=relatively good, 3=relatively bad, and 4=bad.

The overall mean rating was 2.76 for the 12 Japanese participants. The overall mean rating was 2.53 for the 16 English participants. There was a significant difference across language background for good-bad dimension; $F (1, 26) = 9.36$, $p < 0.01$, with the Japanese participants having more bad responses than the English participants. The ratings for the Japanese and the English participants are shown in figure 19 for the good-bad categorization task.

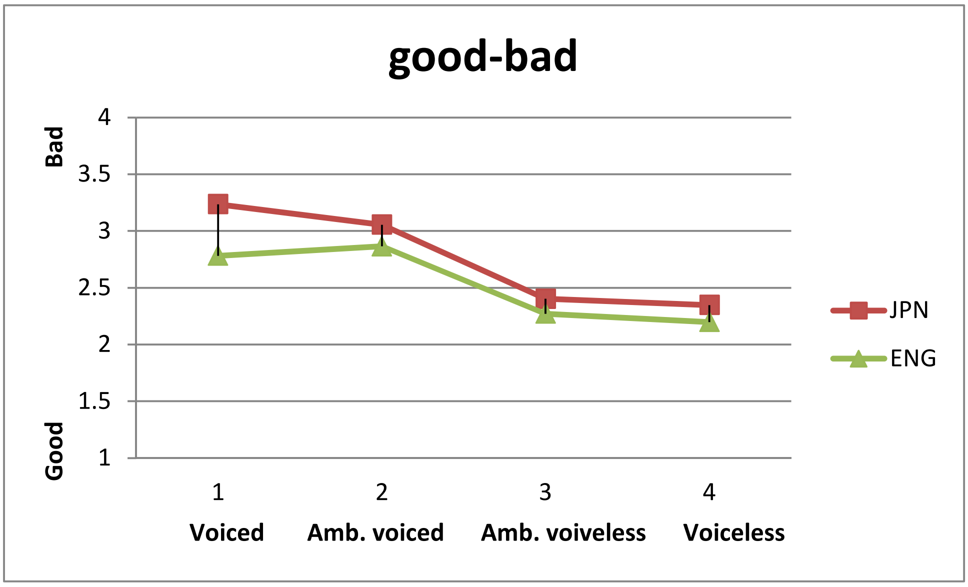

Figure 19: The mean rating for each stimulus for Japanese and English speakers

The mean rating for the 12 Japanese speakers for the voiced stimuli (*deze deze, gede gede, zege zege*) was 3.24, for the ambiguous voiced stimuli 3.06, for the ambiguous voiceless 2.40, and for the voiceless stimuli (*tese tese, kete kete, seke seke*) 2.35.

The mean rating for the 16 English speakers for the voiced stimuli (*deze deze, gede gede, zege zege*) was 2.78, for the ambiguous voiced stimuli 2.86, for the ambiguous voiceless stimuli 2.27, and for the voiceless stimuli (*tese tese, kete kete, seke seke*) 2.20.

There was a significant main effect of voicing; $F(3, 78) = 16.23$, $p<0.001$. Planned comparison tests revealed that significant voicing effects were observed between the voiced and voiceless endpoint stimuli ($p<0.001$). In addition, significant differences were observed between the voiced and the ambiguous voiceless stimuli ($p<0.001$), and between the voiceless and the ambiguous voiced stimuli ($p<0.001$), as well as between the ambiguous voiced and the ambiguous voiceless stimuli ($p<0.005$). There was no significant voicing difference between the voiced and the ambiguous voiced stimuli (p=0.609) and between the ambiguous voiceless and voiceless stimuli (p=0.305).

There was also no overall significant voicing x language interaction; $F(3, 78)= 0.60$, $p=0.620$.

Overall, for the good-bad dimensions, both Japanese speakers and English speakers showed a significant voicing effect but there was no language x voicing interaction. There was a significant language effect which is due to Japanese speakers used more bad responses than the English participants. These results suggest that both Japanese speakers and English speakers associated the voiced (*deze deze, gede gede, zege zege*) and ambiguous voiced stimuli with badness, and they associated the voiceless (*tese tese, kete kete, seke seke*) and ambiguous voiceless stimuli with goodness.

5.4. Graceful-clumsy dimension

For the graceful-clumsy categorization task, participants judged whether the stimulus was graceful or clumsy on a 4-point scale. Numeric values from 1 to 4 were assigned to each member of the scale, with 1=graceful, 2=relatively graceful, 3=relatively clumsy, and 4=clumsy.

The overall mean rating was 2.91 for the 12 Japanese participants. The overall mean rating was 2.52 for the 16 English participants. There was a significant difference across language background for graceful-clumsy dimension; $F(1, 26) = 13.398$, $p < 0.05$, with the Japanese participants overall using more clumsy responses than the English participants. The ratings for the Japanese and the English participants are shown in figure 20 for the graceful-clumsy categorization task.

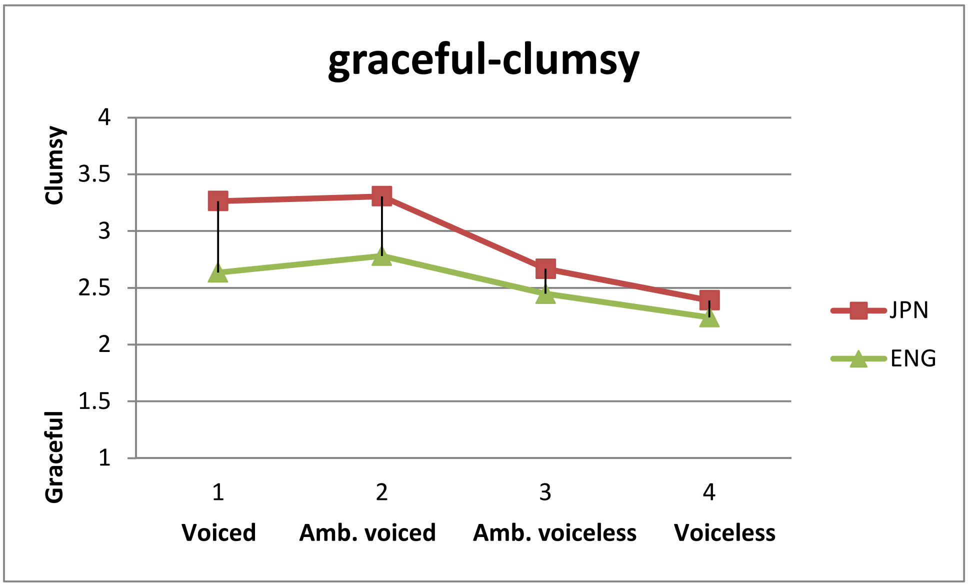

Figure 20: The mean rating for each stimulus for Japanese and English speakers

The mean rating for the 12 Japanese speakers for the voiced stimuli (*deze deze, gede gede, zege zege*) was 3.26, for the ambiguous voiced stimuli 3.31, for the ambiguous voiceless 2.67, and for the voiceless stimuli (*tese tese, kete kete, seke seke*) 2.39.

The mean rating for the 16 English speakers for the voiced stimuli (*deze deze, gede gede, zege zege*) was 2.63, for the ambiguous voiced stimuli 2.79, for the ambiguous voiceless stimuli 2.45, and for the voiceless stimuli (*tese tese, kete kete, seke seke*) 2.24.

There was a significant main effect of voicing; $F(3, 78) = 10.78$, $p < 0.001$. Planned comparison tests revealed that significant voicing effects were observed between the voiced and voiceless endpoint stimuli ($p<0.005$). In addition, significant differences were observed between the voiced and the ambiguous voiceless stimuli ($p<0.005$), and between the voiceless and the ambiguous voiced stimuli ($p<0.001$), as well as between the ambiguous voiced and the ambiguous voiceless stimuli ($p<0.005$). There was no significant voicing difference between the voiced and the ambiguous voiced stimuli ($p=0.281$). However, there was a difference between the ambiguous voiceless and voiceless stimuli ($p=0.018$).

There was no overall significant voicing x language interaction; $F(3, 78)= 1.25$, $p=0.297$.

Overall, for the graceful-clumsy dimension, both Japanese speakers and English speakers showed a significant voicing effect and there was no language x voicing interaction. There was a significant language effect which is due to the Japanese participants overall using more clumsy responses than the English participants. These results suggest that both Japanese speakers and English speakers associated voiced (*deze deze, gede gede, zege zege*) and the ambiguous voiced stimuli with clumsiness, and they associated voiceless stimuli (*tese tese, kete kete, seke seke*) with goodness. There was a voicing effect observed for Japanese speakers between the ambiguous voiceless and voiceless stimuli; Japanese speakers differentiated the ambiguous voiceless stimuli and the voiceless stimuli whereas English speakers categorized both the ambiguous voiceless and voiceless stimuli similarly.

6. Discussion

The current study systematically investigated sound symbolism in voiced consonants /d, g, z/ and in voiceless consonants /t, k, s/ which were embedded in Japanese mimetic non-words. All the non-words were duplicated words (CVCV-CVCV) that consisted of two voiced or voiceless consonants and one vowel /e/ (*deze-deze & tese-tese, gede-gede & kete-kete, zege-zege & seke-seke*). Two subject groups, Japanese native speakers and English native speakers who had no knowledge of Japanese, listened to the non-word stimuli and then rated what they associated the sound with along 4 different dimensions: big-small, round-spiky, good-bad, and graceful-clumsy. Each dimension was rated on a 4-point scale. For the big-small dimension, the 4-point scale ranged from "big", "relatively big", "relatively small", and "small". For the round-spiky dimension, the scale ranged from "round", "relatively round", "relatively spiky", and "spiky". For the good-bad dimension, it ranged from "good", "relatively good", "relatively bad", and "bad". Lastly for the graceful-clumsy dimension, it ranged from "graceful", "relatively graceful", "relatively clumsy", and "clumsy".

The current study showed the existence of sound symbolism in a voicing contrast which is shared by both Japanese and English speakers. Voiced consonants /d, g, z/ tend to be associated with largeness, badness, and clumsiness whereas voiceless consonants /t, k, s/ tend to be associated with smallness, goodness, and gracefulness. Unlike Iwasaki et al. (2007), sound symbolism in the voicing contrast on evaluative dimensions such as good-bad and graceful-clumsy are found in the current study and are found to be shared by Japanese and English participants; both listeners tended to link voiced sounds with badness and clumsiness, and voiceless sounds with goodness and gracefulness.

For the big-small dimension, both Japanese speakers and English speakers showed a significant voicing effect and there was neither a significant language effect nor a language x voicing interaction. Both language groups associated the voiced (*deze deze, gede gede, zege zege*) and less-voiced stimuli with bigness, and they associated the voiceless (*tese tese, kete kete, seke seke*) and less-voiceless stimuli with smallness. There was a clear perceptional distinction between the ambiguous voiced and ambiguous voiceless stimuli for both language groups. Additionally, English speakers showed a slightly greater effect of voicing than Japanese speakers at the ambiguous boundry.

For the good-bad and graceful-clumsy dimensions, both Japanese and English speakers showed a significant voicing effect, a significant language effect, and no significant language x voicing interaction. Both language groups associated the voiced stimuli with badness and clumsiness and the voiceless stimuli with goodness and gracefulness. Interestingly, the distinction between 'bad' and 'good', and between 'clumsy' and 'graceful' on the voiced versus voiceless stimuli was less extreme than that for the big-small distinction. The difference between the voiced and voiceless stimuli on good-bad dimension was 0.89 for Japanese and 0.58 for English speakers, and the difference on graceful-clumsy dimension was 0.87 for Japanese and 0.40 for English speakers. On the other hand, the difference between the voiced and voiceless stimuli on big-small dimension was 0.92 for Japanese speakers and 1.29 for English speakers. Moreover, there was a significant language effect observed for good-bad and graceful-clumsy dimensions. This was due to English participants who provided less 'bad' (more 'good') and 'clumsy' (more 'graceful') ratings than the Japanese participants.

With regard to the sound symbolism on round-spiky shape dimension, no significant effect was found in the voicing contrast in the Japanese participants, but there was a main effect of voicing in the only-stop stimulus pair "*gede-kete*" in the English participants. Kohler (1929) compared the two non-words "baluma" and "takete". He found that English speakers tended to link "baluma" with round shapes and "takete" with spiky shapes. He concluded that continuant strings tend to represent roundness and stop strings tend to represent spikiness. However, there is also a voicing contrast in consonants between "baluma" and "takete"; "baluma" consists of three voiced consonants and "takete" consists of three voiceless consonants. The current study investigated whether the shape symbolism was due to not only the continuant/stop contrast but also due to the voicing contrast. Our overall results showed that this was not the case; there was not a significant voicing effect either for Japanese or English listeners on associating voiced/voiceless sounds with the particular round/spiky dimension. However, the English listeners perceived different images from voiced stops and voiceless stops. They associated voiced stops with roundness and voiceless stops with spikiness. Interestingly, such voicing effect was not observed when the stimuli had both stops and fricatives. In addition, it needs to be noted that there was a trend for Japanese listeners associating voiced sounds with spikiness and English speakers associating voiced sounds with roundness. One possible explanation for not having a significant voicing effect is that two of the three voiced stimuli *"deze deze"* and *"zege zege"* contain both stops and fricatives, and the other stimulus *"gede gede"* consists of only stops. A second possible explanation is that the Japanese participants might have associated the voiced non-word stimuli with an existing mimetic word "*giza giza*" ('sharp' and 'jaggy') which contains two voiced consonants /g/ and /z/. In fact, among the three voiced non-words *"deze deze", "gede gede"* and *"zege zege"*, the one containing both /g/ and /z/, *"zege zege"* was rated to be the most spiky (Japanese mean rating: 3.25), which is followed by the one containing /g/ in the initial consonant position *"gede gede"* (Japanese mean rating: 3.08) and the one containing /z/ in the second consonant position *"deze deze"* (Japanese mean rating: 2.89). Therefore it is possible that the existing Japanese mimetic word "*giza giza*" may have affected the Japanese participants' perception of the stimuli.

The current study created ambiguous continuum for each of the 3 voiced-voiceless stimulus pairs (*deze-deze & tese-tese, gede-gede & kete-kete, zege-zege & seke-seke*) by adjusting the stop and fricative duration of the first and the third consonants (e.g. /d/ in *deze-deze*, /t/ in *tese-tese*, /g/ in *gede-gede,* /k/ in *kete-kete*)*,* and the fricative duration of /z/ in *zege-zege*, and of /s/ in *seke-seke*. This was to investigate categorical perception of voiced sounds to voiceless sounds. One of the two ambiguous members was perceptually closer to the voiceless stimulus and the other was perceptually closer to the voiced stimulus. Therefore, a 4 member continuum for each of the 3 voiceless-voiced pairs was tested.

The ambiguous continua show a distinct categorical perception from voiced to voiceless sounds. The ambiguous stimuli were rated closer to the endpoint on each scale, which indicates that the participants showed strong perceptional distinction between big-small, round-spiky, good-bad, graceful-clumsy when hearing either clearly voiced and voiceless sounds as well as when hearing ambiguously voiced and voiceless sounds.

Mimetic words are commonly and frequently used in Japanese daily lives. However, they are not considered as "official vocabulary" (cannot be found in regular dictionaries). Consequently, the importance of learning mimetic words has not been fully explored, and as a result, they are not taught in most Japanese language classes, nor included in Japanese textbooks. Moreover, they are considered to be one of the hardest word categories to master for adult second language learners (Ivanova 2006). Iwasaki et al (2007) claimed that the difficulty comes from "the obscurity of the image that the mimetic words represent". They found that the sound symbolism in mimetic words is accessible to learners when the word has a high iconicity (i.e. a direct resemblance between the meaning of the word and the sound it refers to; Iwasaki et al. 2007). For instance, mimetic words for laughing have a high iconicity since "they are produced by the same organs as those used to articulate words, the sounds of these words are likely to resemble the quality of voices and manners of laughing that the words refer to" (Iwasaki et al. 2007). In contrast, mimetic words for walking have a low iconicity since "they refer to sounds made by footsteps and to manners that can be perceived visually or by proprioception" (Iwasaki et al. 2007). They concluded that a lack of iconicity of the word can prevent learners from detecting them.

If the difficulty comes from a lack of iconicity, is there any way to fill the void? Can sound symbolism in mimetic words represent the missing iconicity? There are a few studies in which investigated the role of sound symbolism in language learning. Nygaard, Cook, and Namy (2009), Monaghan, Christiansen, and Fitneva (2011), and Monaghan, Mattock, and Walker (2012) are examples of a few studies on the role of sound symbolism in language learning among adults.

Nygaard, Cook, and Namy (2009) examined the contribution of sound symbolism to a novel word-learning task in which monolingual native English speakers were introduced to new Japanese vocabulary. This task was to investigate the influence of sound symbolism on learners' ability to link unfamiliar words to English meanings. The subjects were 104 monolingual native English speakers. They were introduced to 21 Japanese antonym pairs recorded by a native female Japanese speaker. The stimuli were accompanied by English "translations" and grouped into three groups. The first group was the Match condition in which Japanese words were paired with their English equivalents (e.g. *"akarui"* meaning 'bright' paired with an English equivalent 'bright'). The second group was the Opposite condition in which Japanese words were paired with their English equivalent of their antonyms (e.g. *"nageru"* meaning 'throw' paired with an English equivalent of its antonym 'catch'). The third group was the Random condition in which Japanese words were randomly paired with the correct meanings of other words used in the task (e.g. *"akarui"* paired with 'throw').

The results showed a significant main effect on condition; accuracy was significantly better in the Match condition than in the Random condition. However, there was no significant difference between the Match and the Opposite conditions. As for the response time, the pattern across blocks changed as a function of condition. Block 3 showed a significantly better performance in the Match condition than in the Random condition, and no difference was found between the Match and the Opposite conditions. Overall, the subjects responded more quickly in the Match condition in which Japanese words were paired with their correct English equivalents than in the Random condition in which Japanese words were paired with unrelated English words.

Based on these results, Nygaard et al. (2009) suggested that non-arbitrary relationships between sound and meaning exist in natural language and it facilitate on-line acquisition and processing of mappings between sound and meaning. They also claimed that learners are sensitive to this relationship that influences the encoding and retrieval of the meaning of unfamiliar words.

Monaghan et al. (2011) found that sound symbolism facilitate learning categories of a referent. They trained participants to pair pictures of actions and objects with words that varied in phonological features (e.g. plosive vs. non-plosive). The pairings were either in systematic (congruent) or arbitrary (incongruent) relationship between phonological features of words and semantic features of the pictures (e.g. names or shapes). The participants performed well on learning the systematic relationship, thus Monaghan et al. (2011) claimed that there is a correspondence between sound similarity and referent category, in other words, a systematic relationship assisted in learning categories of referents.

Given the finding by Monaghan et al. (2011), Monaghan et al. (2012) investigated whether sound similarity (sound symbolism) assists in learning individual words as well as learning categories. They agreed with the possibility of sound symbolism as an aid to language learning, but they suggested distinguishing learning of categories from learning of individual words. They claimed that “sound symbolism assists learning labels for categories but not for particular referents” (Monaghan et al. 2012). They also claimed that sound symbolism might have effect on language learning, but it is restricted to when category distinctions are important and precise meanings of referents are not important.

Their results indicated that sound symbolism helps learning word-object pairings. However, sound symbolism is useful only when learning word-category relationships and not when learning individual words. They claimed that “sound symbolism assists learning labels for categories but not for particular referents” (Monaghan et al. 2012). They also claimed that sound symbolism might have effect on language learning, but it is restricted to when category distinctions are important and precise meanings of referents are not important.

If sound symbolism assists in learning word-category relationships, it would help learners of Japanese to pair sound symbolic words such as mimetic words with particular category (e.g. shapes, size, and weight). In an initial study, Haryu and Zhao (2007) investigated whether learners of Japanese are more sensitive to sound symbolism. Haryu and Zhao (2007) investigated the symbolic value of voiced sounds and voiceless sounds in Japanese speakers (N=42), Chinese speakers who had studied Japanese (N=40), and Chinese speakers who had no knowledge of Japanese (N=37). In the experiment, the participants were asked to look at two pictures, a small object making a small sound (e.g. a small vase being broken) and a big object making a big sound (e.g. a big vase being broken). They were then asked to listen to either an existing Japanese mimetic word or an onomatopoetic non-word and pick one of the two pictures that matched the sound. There were 14 voiced-voiceless stimulus pairs that consisted of existing Japanese mimetic words and onomatopoetic non-words. The onomatopoetic non-words were created based on 7 existing voiced-voiceless pairs of Japanese mimetic words; the word-initial consonant and the same consonant in the duplicate (either voiced or voiceless) was replaced by another voiced/voiceless consonant in the non-word stimuli.

The results showed that Japanese speakers tended to associate voiced sounds with largeness and voiceless sounds with smallness, whereas Chinese speakers with no knowledge of Japanese did not notice those symbolic values. Most importantly, Chinese speakers who had studied Japanese showed more sensitivity toward the symbolic value than Chinese speakers with no knowledge of Japanese. These results suggest that having experience in the language may play an important role in being aware of the symbolism.

In order to examine whether the amount of experience influences the ability to detect the Japanese-specific sound symbolism, they conducted another experiment with the same stimuli and procedure to compare the performance by second-year learners and fourth-year learners. The results showed no significant difference between the two groups for both existing and novel onomatopoetic words. They claimed that the amount of language experience does not affect learners' ability to detect language-specific sounds symbolism.

In future research, I hope to investigate the learnability of Japanese-specific sound symbolism by examining the perception of voicing in English-speaking learners of Japanese. The current study showed a significant language effect on the two evaluative dimensions, therefore it would be interesting to investigate how learners of Japanese would respond. In order to investigate whether the learnability of Japanese-specific sound symbolism comes from knowledge of existing mimetic words, future research will compare learners' sensitivity to the Japanese-specific sound symbolism between two groups; learners who are knowledgeable of Japanese mimetic words and learners who are not familiar with Japanese mimetic words. If the former group shows higher sensitivity to the Japanese sound symbolism, it would imply that learners may be able to learn language-specific sound symbolism from a certain word category such as mimetic words in Japanese. If there is no significant difference between the two groups, it would suggest that Japanese-specific sound symbolism exists in their entire Japanese lexicon.

Mikami (2007) investigated the recent pedagogical approach to Japanese mimetic words and onomatopoeias. She points out a crucial problem that mimetic words and onomatopoeias are not taught in Japanese language classes, especially in beginner level. This is because mimetic words tend to be considered as supplementary vocabulary; therefore they are not introduced in most of textbooks until advanced levels. However, even in advanced level, only around 30 words are introduced (Mikami, 2007). If the Japanese-specific sound symbolism is learnable and the ability to detect it can be developed by gaining lexical storage of mimetic words, the question arises why should teachers wait to introduce them to learners until they reach intermediate level?

Sound symbolism has been considered to be a marginal phenomenon in linguistics (Imai et al. 2008) since Ferdinand de Saussure's publication stating that the relationship between ‘signifier’ (sound) and ‘signified’ (meaning) is arbitrary. Despite being outside the mainstream, a number of studies have found a relationship between the sound and the meaning of a word. The present study adds some data to address issues of sound symbolism in language. Some sound symbolism in voicing contrast such as size symbolism (voiced-big, voiceless-small), good-bad sound symbolism (voiced-bad, voiceless-good), and graceful-clumsy sound symbolism (voiced-clumsy, voiceless-graceful) were found and demonstrated to be shared across languages.

References

Atouda, T. & Hoshino, K. (2009) Usage guide to Japanese onomatopoeias. *Sotakusha.*

Aveyard M. E. (2012) Some consonants sound curvy: Effects of sound symbolism on object recognition. *Mem Cogn 40, 83–92.*

BMFT Publisher Japan (2012) Words expressing taste.

Hamano, S. (1998) The sound symbolic system of Japanese. *Stanford, CA: CSLI Publications; Tokyo: Kuroshio.*

Haryu, E. & Zhao, L. (2007) Understanding the symbolic values of Japanese onomatopoeia: Comparison of Japanese and Chinese speakers. *Japanese Journal of Psychology, 75, 70-91.*

Hashimoto S., Usui N., Taira M., Nose I., Haji T., & Kojima S., (2006) The neural mechanism associated with the processing of onomatopoeic sounds. *NeuroImage 31, 1762 – 1770.*

Huang, Y.-H., Pratoomraj, S., & Johnson, R. C. (1969) Universal magnitude symbolism. *Journal of Verbal Learning and Verbal Behavior 8, 155-156.*

Iwasaki, N., Vinson, D.P., & Vigliocco, G. (2007) What do English speakers know about gera-gera and yota-yota: A cross-linguistic investigation of mimetic words for laughing and walking. *World's Japanese Language Education 17, 53-78.*

Ivanova J. (2006) Sound-symbolic approach to Japanese mimetic words. *Toronto Working Papers in Linguistics* 26**,** 103–114.

Jorden, E. (1982) Japanese onomatopoeias and mimetic words: Comparison with English. *Nihongo Eigo Hikaku Koza, 111-140.*

Kita, S. (2008) World-view of protolanguage speakers as inferred from semantics of sound symbolic words: A case of Japanese mimetics. *The Origins of Language, 25-38, Springer Japan.*

Kohler, W. (1929) *Gestalt Psychology. New York, USA: Liveright.*

Mikami, K. (2007) Japanese onomatopoeias and the pedagogy. *Dissertation, Waseda University.*

Monaghan P., Mattock K., Walker P. (2012) The Role of Sound Symbolism in Language Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 38, No. 5, 1152–1164.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. Journal of Experimental Psychology: General, 140, 325–347.

Nygaard, L.C., Cook, A.E., & Namy, L.L. (2009) Sound to meaning correspondences facilitate word learning. *Cognition 112, 181-186.*

NHK Close-up Gendai, (June 2013) Japanese onomatopoeias.

Oyama, T., Agostini, T., Kamada, A., Markovic, S., Osaka, E., Sakurai, S., Sarmany-Schuller, I., & Sarris, V. (2008) Similarities in form symbolism among various languages and geographical regions.

Parault S. J. & Parkinson M. (2008) Sound symbolic word learning in the middle grades. *Contemporary Educational Psychology 33, 647-671.*

Sapir, E. (1929) A study in phonetic symbolism. *Journal of Experimental Psychology 12, 225-239.*

de Saussure F. (1983) Course in general linguistics. *R. Harris, trans. Open Court, La Salle, IL (Original work published 1916).*

Shinohara, K. & Kawahara, S. (to appear) A Cross-linguistic Study of Sound Symbolism: The Images of Size. *Tokyo University of Agriculture and Technology* & *Rutgers University.*

Tsujimura N. (2007) An introduction to Japanese linguistics. *Blackwell.*

Ultan, R. (1978) Size-sound symbolism. *J. Greensberg, C. R. Ferguson, & E. A.Moravcsik (Eds.). Universals of human language (Vol. 2, pp. 527-568). Stanford: University Press.*

Watson R. L. (2001) A comparison of some Southeast Asian ideophones with some African ideophones. *Voeltz FKE, Kilian-Hatz C (eds) Ideophones. John Benjamins, Amsterdam, 385-405.*

Westbury C. (2005) Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language 93, 10-19.*

Wrembel, M. (2010) Sound symbolism in foreign language: Phonological acquisition. *Research in Language vol.8.*

Zatorre, R.J., Belin, P., (2001) Spectral and temporal processing in human auditory cortex. *Cereb, Cortex 11, 946–953.*